

Arbeitsbereich Angewandte Sprachwissenschaft und Computerlinguistik, Universität Gießen

Multimedia in der Informationsgesellschaft: Von Open Source zu Open Information

Georg Rehm Henning Lobin

21. März 2000

Inhaltsverzeichnis

1	Einleitung	1
2	Auszeichnungssprachen im World Wide Web	3
3	Das Open Source-Modell in der Software-Entwicklung	7
4	Open Information	11
	Literatur	15

1 Einleitung

Nachdem das Internet bis zum Anfang der neunziger Jahre beinahe ausschließlich von wissenschaftlichen und militärischen Einrichtungen zum Informationsaustausch eingesetzt wurde, hat das intuitiv zugängliche, hypertextbasierte *World Wide Web* (vgl. Berners-Lee, 1999) eine neue Ära des weltumspannenden Computernetzwerks eingeleitet. Waren anfänglich nur vereinzelt Firmen mit Präsenzen im Internet vertreten, so ist es mittlerweile schon fast zu einer Selbstverständlichkeit geworden, dass auch etwa kleinere Handwerksbetriebe die Vorteile der digitalen Medien erkannt haben und mit Homepages im *World Wide Web* für ihre Dienstleistungen und Angebote werben. Neben den zahlreichen Firmen bewegen sich auch immer mehr Schulen – gefördert durch die Initiative „Schulen ans Netz“ des Bundesministeriums für Bildung und Forschung –, gemeinnützige Organisationen, Vereine und Privatanwender im Internet. Die Initiative zur flächendeckenden Versorgung der bundesdeutschen Schulen und Weiterbildungsstätten mit Internet-Verbindungen und zum Einsatz multimedialer Technologien in der Bildung ist eingebettet in das Aktionsprogramm „Innovation und Arbeitsplätze in der Informationsgesellschaft des 21. Jahrhunderts“ der Bundesregierung, dessen Ziel die Sicherung eines Spitzenplatzes in Europa im Bereich der Informations- und Kommunikationstechnologien ist. Das Aktionsprogramm soll u. a. gewährleisten, dass eine beschleunigte Nutzung und Verbreitung dieser Technologien stattfindet, wobei die Nutzung in der Gesamtbevölkerung bis zum Jahr 2005 eine Zielmarke von 30% erreicht haben soll, und dass innovative Arbeitsplätze geschaffen und gefördert werden (Bundesministerium für Bildung und Forschung, 1999). Eine vergleichbare Initiative existiert auch auf europäischer Ebene mit dem Aktionsplan „Europas Weg in die Informationsgesellschaft“ der europäischen Kommission. Der Ursprung dieses Aktionsplans ist ein im Jahr 1994 entstandener Bericht, den eine Gruppe um den damaligen EK-Kommissar Martin Bangemann erarbeitet hat, welcher gezielte Maßnahmen zur Förderung informationstechnologischer Infrastrukturen der Mitgliedsstaaten definiert. Dertouzos (1997, S. 19) fokussiert in diesem Zusammenhang die Frage der Terminologie: „[the] Bangemann Report [...] is a plan for the Global Information Society, as the Europeans like to call the Information Marketplace“. Dertouzos favorisiert letzteren Begriff wegen seiner Ideologiefreiheit und aus einem Verständnis der Informationen als Güter heraus (vgl. Weizenbaum, 1997, S. 34, für eine weiterführende Kritik des Begriffs Informationsgesellschaft), wohingegen in den Vereinigten Staaten vom „Information Superhighway“ oder vom „Cyberspace“, in Japan von einer „National Information Infrastructure“ die Rede ist. Dertouzos kommt zu dem Schluß: „The sudden realization across the globe that the coming world of information will play a key role in people’s lives has caused different nations to put their own imprimatur on [this] new ‘thing’ [...]. The race to coin a name that will prevail is yet another indication of just how big everyone expects the ‘thing’ to be“ Dertouzos (1997, S. 20). Die in verschiedenen Nationen vorgeschlagenen Bezeichnungen für die „Sache“ meinen de facto

das Internet – und hierbei sowohl die technische Infrastruktur als auch Dienste wie die elektronische Post oder das *World Wide Web* und die auf ihm basierenden Angebote – und die Summe der durch den Einsatz eben dieser Dienste zu verzeichnenden und erwarteten Änderungen und Möglichkeiten (siehe hierzu etwa von Haaren und Hensche, 1997): im wirtschaftlichen, gesellschaftlichen und politischen Umfeld, in der Beschäftigungssituation, in der Aus- und Weiterbildung und in der privaten und geschäftlichen Kommunikation, um nur einige der involvierten Gebiete zu nennen; Grundlagen, Definitionen, Formate und psychologische Untersuchungen des Bereichs, der unter dem Schlagwort *Multimedia* fungiert, finden sich – die Literatur zu diesem diffusen Themenkomplex ist geradezu erdrückend – beispielsweise in Pfammatter (1998, S. 9–18), Nielsen (1996), Fluckiger (1996), Steinmetz (1999) und Hasebrook (1995).

Der vorliegende Beitrag betrachtet einen Bereich des multimedialen *Information Marketplace*, der unserer Meinung nach von einer äußerst dynamischen Entwicklung geprägt sein wird und den Benutzern des *World Wide Web* völlig neue Möglichkeiten bei der Informationsrecherche geben wird, die vorwiegend über Suchmaschinen durchgeführt wird. Die momentane Situation im Gebrauch von Suchmaschinen ist vor allem geprägt von einer semantischen Unschärfe, die sowohl aus mangelnder Strukturierung der indexierten Dokumente als auch aus unzureichenden Methoden zur Informationsextraktion resultiert. Konkret bedeutet dies für den Benutzer, dass zu einem Suchbegriff meist viele hundert oder tausend potenziell relevante Dokumente von der Suchmaschine gemeldet werden, eine wirkliche Relevanz zum gegebenen Stichwort aber nicht zwangsläufig vorhanden ist, so dass der Benutzer in mühevoller Arbeit die Liste der Treffer auf ihren tatsächlichen Bezug zur Suchanfrage überprüfen muß.

Dem erwähnten Mangel an Struktur in Web-Dokumenten wird in Zukunft durch den vermehrten Einsatz von XML (*Extensible Markup Language*, Bray et al., 1998) und der gleichzeitigen Aufgabe von HTML (*Hypertext Markup Language*, Raggett et al., 1997), das lediglich eine sehr grobe Auszeichnung textueller Elemente wie Überschriften, Tabellen oder Absätze zuläßt, Einhalt geboten werden; jedoch birgt die neue strukturelle Vielfalt und Freiheit von XML auch Gefahren. Diese liegen unserer Ansicht nach zu einem großen Teil in der immer wiederkehrenden Neuerfindung des Rades: Da XML eine freie Definition konkreter Auszeichnungssprachen wie etwa HTML gestattet, werden viele auf XML basierende Auszeichnungsschemata kreiert werden, die wiederum die suchmaschinenbasierte Recherche in Dokumenten, die unter Benutzung dieser Schemata annotiert wurden, unnötig erschweren, da eine Ausnutzung der Vorteile von XML in einem verteilten Netz wie dem Internet zu einem Großteil auf einer Standardisierung der Auszeichnungsverfahren basiert. Im folgenden skizzieren wir eine Entwicklung, die unserer Ansicht nach dem XML-Babel entgegenwirken könnte. Hierbei geht es um ein seit etwa 20 Jahren erfolgreiches Paradigma in der Software-Entwicklung, das – nicht zuletzt aufgrund des Erfolges des freien Betriebssystems *Linux* – seit 1998 als *Open Source*-Software-Entwicklung in aller Munde ist und der Erschaffung neuer, quasi-standardisierter XML-basierter Auszeichnungssprachen ent-

scheidende Impulse geben wird. Das Resultat dieser Impulse bezeichnen wir als *Open Information*.

Abschnitt 2 führt zunächst in die Thematik ein, indem der Status Quo der heute gegebenen Möglichkeiten zur Auszeichnung von Informationen im *World Wide Web* skizziert wird. Dabei betrachten wir neben XML auch neue, auf diesem Standard basierende Vorschläge zur expliziten Markierung von Metainformationen und zum Aufbau von Konzepthierarchien. Abschnitt 3 erläutert den Ursprung, die Motivationen und das aktuelle Verständnis des Begriffs *Open Source*. Der letzte Abschnitt verknüpft das Paradigma der *Open Source*-Software-Entwicklung mit der kollaborierten Erschaffung und Pflege XML-basierter Auszeichnungssprachen und geht auf die Möglichkeiten ein, die *Open Information* dem Internet und seinen Benutzern geben kann.

2 Auszeichnungssprachen im World Wide Web

Die Auszeichnungssprache des *World Wide Web*, mit deren Hilfe Web-Dokumente geschrieben werden, heißt HTML, *Hypertext Markup Language* (Raggett et al., 1997). HTML gestattet die Anreicherung einer im ASCII-Format vorliegenden Textdatei mit einer klar definierten Menge von Formatierungsmarkierungen, sog. *Tags*; so markiert beispielsweise das *Tag* `<P>` (für *Paragraph*) den Beginn eines Absatzes, `</TABLE>` das Ende einer Tabelle, und Text, der von den *Tags* `<H1>` und `</H1>` (für *Headline*) umrahmt wird, stellt eine Überschrift erster Stufe dar. Weitere *Tags* gestatten die Auszeichnung tieferer Ebenen von Überschriften, verschiedener Arten von Listen (nummerierte und nicht nummerierte, Definitionslisten) und vor allem die Integration von Hyperlinks, also Querverweisen, die den Leser des Textes bei Aktivierung automatisch zu weiterführenden Informationen zu einem bestimmten Begriff führen. Die diesen *Tags* zugrunde liegenden Strukturelemente sind nicht etwa beliebig kombinierbar, sondern es existiert eine regelbasierte, formale Definition, die die Namen und das Zusammenspiel der Elemente spezifiziert. Diese Dokumenttyp-Definition (DTD, *Document Type Definition*) für HTML wurde mit Hilfe der *Standard Generalized Markup Language* (SGML, ISO8879, 1986) definiert, die eine abstrakte und äußerst komplexe Sprache zur Definition konkreter Auszeichnungssprachen (wie eben etwa HTML) darstellt.

Die Ursprünge von HTML befinden sich im europäischen Kernforschungszentrum CERN, wo Tim Berners-Lee und Robert Cailliau 1989 angefangen hatten, ein verteiltes, also im Netzwerk arbeitendes Hypertext-System zu entwickeln, das von den Entwicklern später *World Wide Web* (benannt nach dem von Berners-Lee implementierten ersten grafischen Web-Browser) genannt wurde (Berners-Lee, 1999). Neben einem Protokoll zum Transfer von Hypertext-Dokumenten (HTTP, *Hypertext Transfer Protocol*) und einem Adressierungsschema für beliebige Typen von Web-Dokumenten entwickelten Berners-Lee und Cailliau auch HTML als Formalismus zur Repräsentation von Hypertexten. Die Entscheidung, HTML als „SGML-ähnlich“ (Berners-Lee,

1999, S. 41) zu definieren, war zu einem großen Teil politischer bzw. diplomatischer Natur. SGML wurde zu dieser Zeit am CERN großflächig eingesetzt, so dass Berners-Lee die wichtigsten Elemente des am CERN benutzten SGML-Systems übernahm, um den Mitarbeitern einen leichten Einstieg in HTML zu ermöglichen. Erst einige Jahre später wurde HTML – in einer weiterentwickelten Version – tatsächlich als „echte“ SGML-Anwendung spezifiziert.

Erlaubte HTML anfänglich nur die explizite Auszeichnung sehr grober textueller Elemente wie etwa Abschnitte, Überschriften oder Listen, implementierten die großen Browser-Hersteller – einhergehend mit dem stetigen Erfolg des *World Wide Web* – immer mehr Elemente, die allein auf das Aussehen der HTML-Dokumente am Bildschirm Auswirkungen hatten, so dass der eigentliche Vorteil SGML-basierter Auszeichnungssprachen, die darstellungsunabhängige Explizierung struktureller Information, immer mehr in den Hintergrund rückte. Die Hersteller versuchten damit einerseits, Alleinstellungsmerkmale ihrer Produkte zu definieren, zum anderen aber durch die Schaffung von de facto Standards die langwierigen Entscheidungsprozesse der offiziellen Standardisierungsgremien zu umgehen. Der bereits angesprochene Erfolg des *Web* vor allem im kommerziellen Bereich machte jedoch deutlich, dass die Vorhaltung explizit strukturierter Information ein entscheidendes Kriterium für den Fortbestand des *Web* sein wird, etwa im Bereich des Austausches von aus Datenbanken gespeisten Produktinformationen.

Die explosionsartige Ausbreitung des *World Wide Web* und die Erschließung neuer Interaktionsmöglichkeiten, oftmals im Zusammenhang mit der Entwicklung von *e-Commerce*-Anwendungen, haben schließlich deutlich gemacht, dass für das Internet mit HTML nur ein erster Schritt bei der Informationsmodellierung unternommen worden ist. Da HTML lediglich eine SGML-Anwendung unter anderen ist und andere SGML-Anwendungen auch für die entstehenden Bedürfnisse im *World Wide Web* passender erscheinen, stellt sich die Frage, warum nicht anstatt nur einer bestimmten SGML-Anwendung SGML insgesamt über das Internet nutzbar sein kann. Für bestimmte Anwendungszwecke spezialisierte SGML-Anwendungen könnten dann im Internet verfügbar sein und, darauf aufbauend, auch flankierende Standards wie HyTime (*Hypermedia/Time-Based Structuring Language*, ISO10744, 1997) für weitergehende Verlinkungstechniken und DSSSL (*Document Style Semantics and Specification Language*, ISO10179, 1996) für Strukturtransformationen und flexible Gestaltung mit den dazugehörigen Software-Systemen.

So naheliegend dieser Gedanke ist, so schwierig ist es, ihn umzusetzen. Das *World Wide Web* hat als ein neues Massenmedium längst Fakten geschaffen, die nicht zu dem sehr umfangreichen SGML-Standard passen. Das größte Problem besteht darin, dass der Standard nicht nur kompliziert, sondern auch formal so komplex ist, dass Online-Anwendungen Schwierigkeiten bekommen, ihre Verarbeitung in akzeptabler Zeit durchzuführen. Viele Eigenschaften von SGML spiegeln noch den Stand der frühen achtziger Jahre wider, in denen noch nicht absehbar war, dass SGML-

Anwendungen woanders als auf isolierten Einzelrechnern funktionieren könnten.

Das Hervortreten dieser Unzulänglichkeiten von SGML für Zwecke der Online-Anwendung war der Ursprung von XML, das seit Anfang 1998 in einer vom *World Wide Web Consortium* verabschiedeten Fassung vorliegt (Bray et al., 1998). XML ist nichts anderes als eine vereinfachte Version von SGML – alle in XML kodierte Information ist zugleich auch gültige SGML-Information. Die Definition von XML ist jedoch viel konziser, knapper und logisch überzeugender, da alles das, was in SGML ohnehin kaum genutzt oder heutzutage nicht mehr gebraucht wird, weggelassen wurde, ohne dabei die Ausdrucksmöglichkeiten prinzipiell einzuschränken. Diese Reduktion ist so überzeugend gelungen, dass XML inzwischen auch dort eingesetzt wird, wo die Online-Fähigkeit der Daten gar nicht im Vordergrund steht.

Man kann XML als ein Instrument für die Modellierung von strukturierter Information verstehen (Lobin, 2000). Was ist strukturierte Information? Der Idee der strukturierten Information liegen verschiedene Beobachtungen zugrunde, die ursprünglich an Textdokumenten gemacht worden waren: In einem Text können erstens unterschiedliche Ebenen voneinander unterschieden werden. Es gibt einerseits die Abfolge von Buchstaben, z. B. in einer Überschrift oder als ein Zitat, es gibt andererseits aber auch abstrakte Einheiten, die z. B. für die Kategorien „Überschrift“ oder „Kapitel“ stehen. Diese abstrakten Einheiten werden im Gegensatz zu den textuellen Einheiten nicht durch sprachliche Zeichen konkretisiert, sondern oftmals durch typografische: die Schrift einer Überschrift ist größer als die des folgenden Textes, die Überschrift wird abgesetzt und meistens nummeriert, und auch für die Kennzeichnung eines zusammenhängenden Textteils als ein Kapitel gibt es verschiedene Darstellungsmittel.

Die zweite Beobachtung ist, dass die Anordnung der abstrakten und der konkreten Informationseinheiten nicht beliebig ist, sondern vielmehr festen Regeln zu folgen hat, die denen zur Bildung von Sätzen ähneln. Diese Regeln spezifizieren einerseits das hierarchische Verhältnis von abstrakten Informationseinheiten zu untergeordneten abstrakten oder konkreten Informationseinheiten, andererseits die lineare Abfolge gleichrangiger Informationseinheiten. Man kann diese Regeln zu einer Grammatik der Informationseinheiten – der bereits angesprochenen *Document Type Definition* – zusammenfassen.

Die dritte Beobachtung: Eine solche Grammatik kann immer so gestaltet werden, dass sich die Informationseinheiten mit ihren hierarchischen und linearen Beziehungen zueinander in Baumform anordnen: ganz oben gibt es ein Wurzelement, das den Text als Ganzes repräsentiert, die Töchter darunter repräsentieren die Teile, aus denen sich der Text auf oberer Ebene zusammensetzt, und diese Zerteilung wird solange fortgesetzt, bis man auf der Ebene der elementaren Texteinheiten angelangt ist.

In XML sind diese Beobachtungen in einen systematischen, formal definierten Zusammenhang gebracht worden:

- Eine XML-Anwendung gibt an, was für Typen von abstrakten und konkreten

Informationseinheiten es gibt, gibt ihnen Namen zur eindeutigen Identifizierung und spezifiziert gegebenenfalls weitere Beschreibungsmerkmale.

- Diese Informationstypen werden durch Regeln miteinander in Beziehung gesetzt.
- Diese Regeln werden zu einer Grammatik zusammengefasst.
- Reale Informationseinheiten werden mit diesen Typen in Beziehung gesetzt und in Baumform angeordnet.

Strukturierte Information ist also nichts anderes als die regelgeleitete Anordnung von Informationseinheiten, genauso wie wir korrekt strukturierte Sätze als regelgeleitete Anordnung von Wörtern verstehen können.

Parallel zu XML vom *World Wide Web Consortium* entwickelte Formalismen beschäftigen sich mit der Verknüpfung von XML-Dokumenten (XPointer, XLink), der Verarbeitung mehrerer Dokumenttypdefinitionen in einem Dokument (*Namespaces*) und der Visualisierung von XML-Dokumenten in Web-Browsern (*Cascading Style Sheets*, CSS, und *Extensible Style Language*, XSL).

Mit der Einführung von XML ist beim Umgang mit Information ein entscheidender Schritt vollzogen worden: erstmals wird es möglich, Informationen nicht nur aus einer technologischen Perspektive zu betrachten, sondern auch aus einer inhaltlichen. XML-strukturierte Daten sind unabhängig von bestimmten Software-Systemen oder gar Betriebssystemen, sie sind im Normalfall sogar unabhängig von der Darstellung in einem bestimmten Medium. Stattdessen geben sie Auskunft über ihren logischen Aufbau und vermögen im Idealfall diesen Aufbau aus den semantischen Eigenschaften der Daten abzuleiten. Sind diese Möglichkeiten im *World Wide Web* nur in sehr eingeschränkter Form genutzt worden, so ist für die nächste Zeit damit zu rechnen, dass inhaltsorientierte Navigations-, Such und Präsentationsverfahren an Bedeutung gewinnen.

Ein vom *World Wide Web Consortium* vorgeschlagenes Verfahren zur Schaffung einer Basis, die eine gezielte und sinnvolle Wissensexploration ermöglichen soll, ist das XML-basierte *Resource Description Framework* (RDF, siehe Lassila und Swick, 1999; Brickley und Guha, 1999). RDF gestattet die Auszeichnung von Metadaten, Daten über Daten, durch den Autor oder Bearbeiter eines Web-Dokuments. Beispiele für Metadaten eines Web-Dokuments sind etwa der Name des Autors, das Datum der letzten Änderung, verschiedene Schlagworte, ein Verweis auf die assoziierte Organisation etc. RDF wurde entwickelt, um eine umfassende und konsistente Explizierung von Metadaten in Web-Dokumenten zwecks vereinfachter und präziserer Suche und Exploration von Dokumentbeständen zu gewährleisten. RDF gestattet, ähnlich wie XML, lediglich die Definition verschiedener Schemata, mit deren Hilfe dann wiederum konkrete Dokumente annotiert werden können. Hierbei ergeben sich einige unmittelbare Probleme: Welche (standardisierten?) Vokabulare werden zur Definition von RDF-Schemata eingesetzt? Wie detailliert sollen die Metadaten strukturiert werden und welche Arten

von Metadaten sollen – jeweils abhängig vom Themengebiet – annotierbar sein? Zu diesen Fragestellungen der generellen Klassifikation von Objekten gibt es in verschiedenen Fachrichtungen (Bibliothekswesen, Architektur, Kunst etc.) Bemühungen zur Schaffung von Standards (Übersichten befinden sich etwa in Hudgins et al., 1999; Baca, 1998; Marchiori, 1998); im *World Wide Web* scheint sich mehr und mehr die noch in der Entwicklung befindliche Initiative *Dublin Core* (siehe <http://purl.org/dc/> und Weibel et al., 1999) durchzusetzen, die eine Art gemeinsamen und erweiterbaren Kern aller RDF-Schemata hervorbringen soll. Das *Dublin Core* Schema definiert drei verschiedene Gruppen von Elementen: *Content* (mit Elementen wie etwa *Title*, *Subject*, *Description*, *Type* etc.), *Intellectual Property* (*Creator*, *Publisher*, *Contributor*, *Rights*) und *Instantiation* (*Date*, *Format* etc.). RDF wird – gerade in Verbindung mit dem *Dublin Core* – mittlerweile schon auf vielen Websites intern eingesetzt, um Ressourcen, also Web-Dokumente und in ihnen eingebettete Objekte, zu beschreiben.

Mit *Topic Maps* (ISO/IEC13250, 1999, vgl. Rath, 1999, früher *Topic Navigation Maps*) liegt bereits eine XML-Architektur für die standardisierte Darstellung von Metainformation vor. Dabei geht es – im Gegensatz zu RDF – vor allem um die thematischen Bezüge zwischen Informationsobjekten. Zweck einer solchen Darstellung ist die Unterstützung von inhaltsorientierter Navigation und Filterung. Die grundlegende Idee besteht darin, beliebige Informationsobjekte zu Gruppen zusammenfassen und auf einer abstrakten Ebene zu gliedern. Der Bezug zu den konkreten Informationseinheiten kann dabei durch Verweise hergestellt werden, die Gesamtheit der thematischen Bezüge wird allerdings auch allein als ein sinnvoller unabhängiger Informationstyp verstanden. Die Bezüge zwischen Informationseinheiten, die sog. *associations*, können ihrerseits zu Gruppen geordnet werden, so dass auch Filterungen auf dieser Ebene möglich werden.

Der Vorteil der Standardisierung von thematischen Strukturierungen ist darin zu sehen, dass für unterschiedliche Domänen oder Wissensbereiche erstellte *Topic Maps* verschmolzen werden können und so nach einiger Zeit ein allumfassendes begriffliches Netzwerk entstehen kann. Es ist denkbar, dass das *World Wide Web* in seiner heutigen Erscheinungsform oder andere Hypertext-Systeme um ein Netz von begrifflichen Bezügen angereichert werden, die zusätzlich zu der Nutzung explizit repräsentierter Verlinkungen die Exploration zusammenhängender Wissensbereiche erlauben.

3 Das Open Source-Modell in der Software-Entwicklung

Unter dem *Open Source*-Modell in der Software-Entwicklung (siehe Vixie, 1999, für einen Vergleich dieses Ansatzes mit traditionellen Methoden) versteht man die freie Veröffentlichung, Weitergabe und Erlaubnis zur Modifikation von Programm-Quellen – den in Programmiersprachen wie beispielsweise *C* oder *Java* geschriebenen Instruktionen, die vor der Ausführung durch den Computer in Maschinenbefehle konvertiert werden müssen (vgl. DiBona et al., 1999). Der Erfolg dieses Ansatzes (die meisten

Email- und *World Wide Web*-Server werden mit *Open Source*-Entwicklungen betrieben, vgl. O'Reilly, 1999) wird als ein Phänomen betrachtet, das in engem Zusammenhang mit dem Internet steht und umfangreiche Auswirkungen auf die Branche der Kommunikations- und Informationstechnologie hatte und auch in Zukunft haben wird. Im folgenden erläutern wir die Entstehung des *Open Source*-Ansatzes, nennen die wichtigsten Gründe für den Erfolg von *Open Source*-Software und gehen auf Erweiterungen dieses Paradigmas ein.

Der heutige *Open Source*-Begriff geht auf Gedanken zurück, die erstmalig Mitte der achtziger Jahre von dem Software-Entwickler Richard M. Stallman geäußert wurden. Stallman hat zu dieser Zeit seine Anstellung im Labor für Künstliche Intelligenz des Massachusetts Institute of Technology aufgegeben, um sich ganz der Entwicklung eines freien – “free as in freedom” (Stallman, 1999), also nicht *frei* im Sinne von *kostenlos* – Betriebssystems zu widmen, das zum De-facto-Standard der Industrie, dem UNIX-System, kompatibel sein sollte. Stallman war nicht mit der allgemeinen Tendenz vieler Software-Anbieter einverstanden, die Weitergabe der Quell-Dateien ihrer Produkte an die Benutzer abzulehnen, da diese die Quellen als schützenswertes Eigentum verstanden haben. Quell-Dateien ermöglichen es dem geübten Anwender und Programmierer, Fehler in Programmen zu suchen und zu beheben, neue Funktionen in bestehende Programme zu integrieren oder aus Teilen verschiedener Programme und einem gewissen Maß an Eigenentwicklung Software mit ganz neuer Funktionalität zu erschaffen. Ohne die Programmquellen haben Anwender diese Möglichkeiten nicht, sondern können etwa im Falle eines Programmfehlers nur hoffen, dass dieser in der nächsten Version der eingesetzten Software behoben sein wird. Stallman war der Ansicht, dass Programmierer ein ethisches Anrecht auf freie Software haben und begann mit der Arbeit an dem Betriebssystem GNU (dessen rekursiv definiertes Akronym “GNU’s not UNIX” bedeutet). Auf der Grundlage eines bestehenden, kommerziellen UNIX-Systems implementierte Stallman seine Versionen eines Editors (*Emacs*), eines Compilers (*gcc*) und verschiedener Werkzeuge (*gdb*, *make*), um die Standardkomponenten des von ihm eingesetzten, kostenpflichtigen UNIX-Systems sukzessive durch freie, eigenentwickelte Module zu ersetzen. Weiterhin nahm er, soweit dies möglich war, bereits verfügbare freie Software wie das Satzsystem *TEX* oder das Fenstersystem *X Window* in das GNU System auf (Buthenuth und Mock, 1992). Mit der Entwicklung des vielfältig konfigurierbaren Editors *Emacs* wuchs sowohl das allgemeine Interesse an der Philosophie der freien Software, als auch das Interesse vieler Programmierer an einer Mitarbeit an verschiedenen GNU-Komponenten. Stallman gründete daraufhin die *Free Software Foundation* (FSF), eine gemeinnützige Vereinigung zur Verbreitung und Unterstützung seines Ziels, ein im Quelltext vorliegendes UNIX-kompatibles Betriebssystem zu erschaffen, im Zuge dessen auf Probleme bezüglich Urheberrecht, Modifikation und Verkauf von Software aufmerksam zu machen sowie diese Probleme gänzlich zu „eliminieren“ (die FSF ist im *World Wide Web* erreichbar unter <http://www.fsf.org>).

Als die ersten Komponenten des GNU Systems zu ausgereifter Software herange-

wachsen waren, benötigte Stallman eine rechtliche Absicherung, um eine Übernahme der von ihm entwickelten freien Software durch Firmen auszuschließen, die seine Quell-dateien evtl. modifizieren und daraufhin als proprietäres Firmeneigentum ausweisen könnten. Unter anderem aus dieser Motivation (siehe Stallman, 1999, für weitere Beweggründe) entstand die GNU *General Public License* (GPL) und die Idee des “Copyleft” (“All rights reversed”) Vertriebskonzepts. Der Begriff *Copyleft* meint hier die Wahrung der Urheberrechte des Autors (“Copyright”) an der von ihm entwickelten Software zu genau dem Zweck, diese als freie Software auszuweisen. Diese Lizenz gibt also jedem das Recht, unter der GPL lizenzierte Software zu benutzen, beliebig oft zu kopieren und zu modifizieren und auch modifizierte Versionen zu vertreiben. Jedoch dürfen der Software, modifiziert oder nicht, keinerlei Restriktionen auferlegt werden. Mit diesen Lizenzbestimmungen erklärt man sich einverstanden, indem man die entsprechende Software benutzt, erweitert oder vertreibt (§5 der GPL). Des weiteren unterliegt jedes Programm, das von GNU/GPL Software abgeleitet wird, zwingend wieder der GPL, so dass mit Hilfe des Copyleft Modells garantiert werden kann, dass ursprünglich freie Software auch in alternativen oder erweiterten Versionen in Zukunft den Status freier Software im Sinne der GPL/FSF besitzt.

Im Laufe einiger Jahre entstand im Rahmen der Initiative und Koordination der *Free Software Foundation* eine Sammlung vielzähliger Programme. Häufig beteiligten sich dutzende Entwickler an der Programmierung eines Software-Pakets, oftmals koordiniert über das Internet, mit dessen Hilfe auch neue Versionen einzelner Module ausgetauscht wurden. Das GNU System konnte jedoch noch nicht als vollständiges und vor allem eigenständiges UNIX-System fungieren, da die wichtigste Komponente noch fehlte: der Kern. Der Kern eines Betriebssystems ist verantwortlich für die Speicher- und Prozeßverwaltung, Zugriffe auf die Hardware etc. Diese Lücke wurde mit der Entwicklung von *Linux* gefüllt. *Linux* bezeichnet heutzutage ein nicht nur in Forschung und Lehre weit verbreitetes UNIX-kompatibles Betriebssystem, das gegen einen geringen Unkostenfaktor auf Datenträgern bzw. kostenlos im Internet zur Verfügung steht. Der Begriff *Linux* meint jedoch ursprünglich lediglich einen UNIX-Kern, entwickelt von dem finnischen Informatik-Studenten Linus Torvalds. Torvalds konnte Anfang der neunziger Jahre mit Hilfe der sehr effektiven Kommunikationsmöglichkeiten des Internet in kurzer Zeit viele begeisterte Anhänger für sein komplexes Vorhaben, dessen Entwicklungsbasis das zu Lehrzwecken entwickelte UNIX-System *Minix* war, finden (Raymond, 1999, berichtet, dass bereits von Dritten implementierte Ansätze für ein Programm ein wichtiger Faktor für den Erfolg eines *Open Source*-Projekts sind). Er veröffentlichte sehr oft – durchschnittlich etwa einmal pro Woche – neue Versionen des Kerns, bekam innerhalb weniger Tage von hunderten freiwilliger Mitentwickler per elektronischer Post Rückmeldungen über Fehler der neuesten Version und integrierte ebenfalls mitgeschickte Erweiterungen und Verbesserungsvorschläge in die Quell-Dateien. Auf diese Weise entstand in bemerkenswert kurzer Zeit ein voll funktionsfähiger und leistungsstarker UNIX-Kern, der auf einem handelsüblichen

PC lauffähig war und somit in idealer Weise die problematischste Lücke der GNU-Programmsammlung schließen konnte. Dadurch ist das *Linux*-System in der heute bekannten Form entstanden und wird noch immer von vielen tausend Programmierern weltweit – koordiniert mit Hilfe der Kommunikationsmöglichkeiten des Internet – weiterentwickelt.

Eine Gruppe um den Programmierer Eric S. Raymond hat Anfang 1998 die Potenziale des Ansatzes erkannt, der zur Entwicklung so effizienter und leistungsfähiger freier Software wie *Linux*, dem De-facto-Standard der *Email-Server (Sendmail)*, dem sehr erfolgreichen *World Wide Web-Server Apache* oder der Programmiersprache *Perl* geführt hat: begeisterte Entwickler sind – verbunden durch das Internet – in der Lage, nur aufgrund ihres eigenen Ehrgeizes, der Freude am Programmieren und der Qualifikation für einen späteren Beruf leistungsstarke, sichere und stabile Software in einer Geschwindigkeit zu implementieren, die in herkömmlichen Entwicklungsabteilungen von Software-Firmen niemals möglich wäre (Ettrich, 2000). Wenn ein Benutzer eines Software-Pakets beispielsweise mittels elektronischer Post einen Fehler meldet, ist es ihm des weiteren – Programmierkenntnisse vorausgesetzt – aufgrund der offenliegenden Quellen möglich, unmittelbar einen Verbesserungsvorschlag an die federführenden Entwickler zu schicken, so dass bereits nach kurzer Zeit eine neue, verbesserte Version des Programms über das Internet verbreitet werden kann (Raymond, 1999). Die Gruppe um Raymond nannte dieses Prinzip “*Open Source Software*” (siehe <http://www.opensource.org>), um sich von dem potenziell mißverständlichen Begriff der freien Software abzugrenzen, und sie versteht *Open Source* als Marketing-Strategie, um das erwiesenermaßen erfolgreiche Konzept der freien Software kommerziellen Software-Herstellern näher zu bringen. Zwischen den beiden Gruppierungen, der *Free Software Foundation* um Stallman auf der einen, und der *Open Source Initiative* um Raymond auf der anderen Seite, herrschen gewisse Differenzen (Stallman spricht von „zwei politischen Parteien innerhalb einer Gemeinschaft“), da erstere vor allem den sehr idealistischen Aspekt der Freiheit von Software, jedoch letztere nur die effektive Entwicklung leistungsstarker Software in den Vordergrund stellt (siehe hierzu etwa Feuerbach und Schmitz, 1999).

Seit einiger Zeit zeichnet sich ein Trend in der freien Software-Szene ab, sich nicht länger auf Betriebssystem-spezifische Grundlagen und ihre Implementation zu beschränken, sondern vermehrt auch Anwendungs-Software zu entwickeln, etwa die komfortable grafische Arbeitsoberfläche KDE (*K Desktop Environment*) und die hierin eingebettete Applikations-Sammlung *KOffice*, die leistungsstarke Werkzeuge zur Textverarbeitung oder zur Kalkulation beinhaltet. Aber auch rudimentäre Forschungsprototypen wie beispielsweise die Spracherkennung *Sphinx* der Carnegie Mellon University sind mittlerweile unter der *Open Source Lizenz* frei verfügbar, und somit ist absehbar, dass mittelfristig – freiwillige Programmierer vorausgesetzt – eine ausgereifte Komponente zur Spracherkennung basierend auf *Sphinx* unter *Linux* einsetzbar sein wird.

Das *Open Source*-Paradigma ist mittlerweile nicht mehr ausschließlich auf den Bereich der Software-Entwicklung beschränkt. Anfang 1998 wollte David Wiley, zu dieser Zeit Doktorand an der Brigham Young University, verschiedene Lehrmaterialien im *World Wide Web* veröffentlichen, aber gleichzeitig sicherstellen, dass zum einen im Falle einer Verwendung sein Name genannt wird, zum anderen die von ihm erstellten Materialien nicht entgegen ihrem Zweck verändert werden. Mit der Unterstützung von Stallman und Raymond hat Wiley auf der Grundlage der GNU *General Public License* die *Open Content License* (OPL) entwickelt (siehe <http://www.opencontent.org>). Diese gestattet eine kostenfreie Verwendung von Inhalten und erzwingt eine unmißverständliche Markierung derjenigen Stellen, die von Dritten modifiziert werden. Im Falle einer Veränderung sieht diese Lizenz ebenfalls vor, dass die neuen Inhalte ebenfalls der OPL unterliegen. Auf der Website der *Open Content Initiative* werden zum Zeitpunkt des Verfassens dieses Beitrags mehr als 150 Online-Angebote aufgelistet, die unter der OPL veröffentlicht wurden (eine große Suchmaschine findet derzeit zum Stichwort „opencontent“ mehr als 5,500 Treffer). Hierunter befinden sich nicht nur Lehrmaterialien zu den verschiedensten Themen, sondern unter anderem auch Musikstücke und Essays. Weiterhin sind bereits verschiedene Bücher in Planung, die unter einer für die Veröffentlichung von Printmedien modifizierten Fassung der *Open Content License* veröffentlicht werden sollen.

Das Berkman Center for Internet and Society – angesiedelt an der juristischen Fakultät der University of Harvard – untersucht in Projekten wie beispielsweise *Open Law*, *Open Governance* oder *Open Education* „real and possible boundaries in cyberspace between open and closed systems of code, of commerce, of government, and of education, and the relationship of law to each“ (vgl. <http://cyber.law.harvard.edu>). Das Projekt *Open Law* etwa bietet interessierten Internet-Benutzern eine Plattform zum offenen und freien Gedankenaustausch zu aktuellen und Netz-relevanten Fällen, die derzeit vor amerikanischen Gerichten ausgetragen werden. Im Zuge dieser Diskussionen sollen Argumente gefunden und Plädoyers entwickelt werden, wobei eine Prämisse des Projekts ist: „an open development process best harnesses the distributed resources of the Internet community. By using the Internet, we hope to enable the public interest to speak as loudly as the interests of corporations.“

4 Open Information

Die Abschnitte 2 und 3 betrachten zum einen auf der technologischen Ebene den Fortschritt des *World Wide Web* seit dessen Anfängen und zum anderen den Ursprung, die Motivation und Erfolgsfaktoren des *Open Source*-Paradigmas in der Software-Entwicklung. Obwohl diese Themenbereiche zunächst keine wesentlichen Affinitäten aufweisen, könnte sich unserer Meinung nach im Laufe der nächsten Jahre eine interessante Entwicklung vollziehen, die vor allem die Anwendung derjenigen Strategien und Faktoren, die den *Open Source*-Ansatz erfolgreich gemacht haben, auf die Entstehung

von WWW-Inhalten und insbesondere Dokumenten-übergreifende Metastrukturen betrifft.

Abschnitt 2 macht deutlich, dass eine zunehmende Strukturierung des *World Wide Web* stattfindet. Am Beginn stand die erste Fassung von HTML, die Tim Berners-Lee, der Entwickler des WWW, lediglich aus Gründen der internen Vermarktung seiner Idee des *vernetzten* Hypertextes an seiner damaligen Wirkungsstätte als „SGML-ähnlich“ (Berners-Lee, 1999, S. 41 f.) entwickelt hat. In den folgenden Jahren entstanden daraufhin – definiert durch SGML – immer komplexere Versionen von HTML, die immer mehr den Aspekt der Gestaltung und des Layouts von WWW-Dokumenten betonten. Schließlich wurde diese Entwicklung, die aufgrund einer fehlenden Strukturierung der Dokumente in einem Kollaps des *World Wide Web* zu enden drohte, mit der Spezifikation von XML durchbrochen. Nachdem mit XML und den flankierenden Standards (XSL, XPointer, XLink etc.) eine wohldefinierte Strukturiertheit von WWW-Dokumenten gewährleistet sein wird, beschäftigt sich die aktuelle Forschung vornehmlich mit verschiedenen Definitionen und Repräsentationsformalissen für Metadaten, wie beispielsweise dem *Resource Description Framework* (RDF) oder *Topic Maps*.

Die Formalismen RDF und *Topic Maps* stellen unterschiedlich komplexe Möglichkeiten bereit, Informationen über einzelne Web-Dokumente und Zusammenhänge zwischen mehreren Web-Dokumenten zu explizieren. Es werden jedoch auch abstraktere Einsatzgebiete verfolgt: Das *Open Directory Project* (<http://dmoz.org>) stellt einen Web-Katalog (ähnlich dem bekannten Dienst Yahoo) dar, der von etwa 20.000 freiwilligen Internet-Benutzern gepflegt wird, die sich jeweils um einzelne oder mehrere Kategorien kümmern. Diese Pflege umfasst die Aufnahme konkreter Web-Dokumente in bestehende Kategorien der *Open Directory*-Hierarchie und die Schaffung neuer Kategorien, sofern dies erforderlich ist. Die *Open Directory*-Hierarchie ist im graphentheoretischen Sinn als eine Menge von derzeit 15 Bäumen organisiert, die auf der obersten Ebene Konzepte wie *Arts*, *Health*, *Recreation* oder *Science* umfassen und daraufhin in immer speziellere Knoten verzweigen (wie etwa *Arts: Movies: Genres: Silent Movies* oder *Science: Social Sciences: Language and Linguistics: Applied Linguistics*). Von besonderem Interesse ist hierbei zum einen die Tatsache, dass die Erstellung – ebenso wie das in diesem Umfeld bislang unbekannte Peer Reviewing – des Kataloginhalts durch Benutzer des Angebots und nicht durch eine Vielzahl bezahlter Redakteure wie etwa bei Yahoo vorgenommen wird, und zum anderen, dass die vollständige Hierarchie des *Open Directory Projects* als (von der zugrundeliegenden Datenbank exportierte) RDF-Datei vorliegt und dabei einer GPL-ähnlichen Lizenz unterliegt. Zum Zeitpunkt des Verfassens dieses Beitrags enthält die Hierarchie etwa 200.000 Einträge und die verschiedensten Verknüpfungen innerhalb dieser Struktur.

Betrachtet man nun unter den geschilderten Voraussetzungen die These, dass einer der Faktoren eines erfolgreichen *Open Source*-Projekts eine relevante und modifizierbare Ausgangsbasis (etwa von Dritten implementierte, erste Ansätze zur Lö-

sung eines Problems) ist (Raymond, 1999, S. 33 f.), so bestünde mit der *Open Directory*-Hierarchie eine Grundlage zur Entwicklung einer noch spezielleren Hierarchie bzw. Teilhierarchie im Rahmen von *Open Source*-ähnlichen Projekten. Eine solche Hierarchie – gewissermassen ein semantisches Netz –, beispielsweise aus dem Bereich der Architektur, müsste nicht zwangsläufig als ein exklusiver Teil des *Open Directory Projects* angesehen werden, sondern könnte in den unterschiedlichsten Kontexten, wie etwa in der Computerlinguistik zur Disambiguierung oder auch im Bereich der Entwicklung intelligenter grafischer Benutzeroberflächen oder Online-Navigationssysteme eingesetzt werden. Es ginge nun folglich nicht mehr ausschließlich um die Programmierung von Software (*Open Source*) oder um die Schaffung bzw. Diskussion konkreter Inhalte (*Open Content, Open Law*), sondern um übergreifende und abstrakte Meta-Strukturen, die wir aufgrund des ihnen potenziell inhärenten breiten Anwendungsspektrums als *Open Information* bezeichnen möchten.

Raymond (1999) nennt, wie bereits in Abschnitt 3 angesprochen, weitere Faktoren für den Erfolg des *Open Source* Paradigmas. Dessen sinnvolle Übertragung auf das Konzept der *Open Information* macht eine Überprüfung dieser Faktoren erforderlich. *Open Source* ist erfolgreich, weil begeisterte und erfahrene Programmierer Spaß daran haben, qualitativ hochwertige Software zu entwickeln, um etwa einen persönlichen Bedarf zu befriedigen oder ihre Chancen und Möglichkeiten im Beruf zu erhöhen. Damit diese Faktoren auch für das *Open Information*-Modell gelten können, müssen verschiedene Bedingungen erfüllt sein: Zum einen muß ein Bedarf für frei verfügbare XML- oder RDF-Schemata oder *Topic Map*-Hierarchien bestehen. Dieser Bedarf könnte beispielsweise durch eine großflächige Unterstützung (die bislang noch nicht existiert) dieser Formalismen in den gängigen Web-Browsern, -Editoren und auch dedizierter Software geschaffen werden (Berners-Lee, 1999, S. 172, spricht in einem ganz ähnlichen Kontext von einem “common new genre on the Web”). Sobald Benutzer die Vorteile dieser Technologien, vor allem die vereinfachte Suche und Navigation, erkennen und schätzen lernen, werden sie – das hat die Vergangenheit (von Yahoo zum *Open Directory Project*, von proprietärer Suchmaschinen-Technologie zu *Open Source*-Produkten mit ähnlicher Funktionalität etc.) gezeigt – bestrebt sein, ähnliche Informations-Infrastrukturen zu schaffen und diese als *Open Information* anbieten. Ein weiterer aktueller Trend in der Informationstechnologie ist, dass aufgrund des Erfolges und der Popularität des *Open Source*-Ansatzes immer mehr Firmen die Quelldateien ihrer Software veröffentlichen, um beispielsweise ein globales Peer Reviewing durch die vielen zehntausend Programmierer weltweit zu gewährleisten. Wir halten es für sehr wahrscheinlich, dass es nicht nur bei Software oder Web-Inhalten bleiben wird, sondern dass auch Metastrukturen veröffentlicht werden, die dann unter anderem von Freiwilligen weiterentwickelt und gepflegt werden.

Bei dieser Entwicklung werden sich unserer Meinung nach, vergleichbar den verschiedenen *Open Source*-Projekten, Benutzergruppen aus den unterschiedlichsten Bereichen zusammenfinden, um etwa mit Hilfe von RDF und/oder *Topic Maps* definitive,

den aktuellen Stand des Wissens beschreibende semantische Netze für ihren jeweiligen Interessensbereich zu erstellen. Diese Gruppen werden über das Internet sowohl kommunizieren als auch ihre Projekte und Ergebnisse koordinieren und publizieren. Thematisch benachbarte *Open Information*-Projekte werden dabei versuchen, ihre Hierarchien zu kombinieren und hierbei einheitliche Schemata einzusetzen. Auf diese Weise könnte im Laufe der kommenden Jahre tatsächlich entstehen, was Berners-Lee (1999, S. 177 f.) als das "Semantic Web" ("a web of data that can be processed directly or indirectly by machines") bezeichnet: Die unterschiedlichsten Hypertext-Dokumente, Ton- und Bilddateien des *World Wide Web* werden von den Benutzern selbst in einen globalen Zusammenhang integriert ("This all works only if each person makes links as he or she browses, so writing, link creation, and browsing must be totally integrated", Berners-Lee, 1999, S. 201). Dieser übergreifende Zusammenhang ist, und das ist im Gegensatz zur heutigen Situation der Ungeordnetheit der große Vorteil, *explizit strukturiert*, so dass eine reibungslose maschinelle Verarbeitung, beispielsweise das automatische Schließen zum Zwecke der Erkennung neuer Relationen zwischen nicht unmittelbar benachbarten Gebieten mittels einer Vielzahl thematischer Hierarchien, gewährleistet ist: "We will solve large analytical problems by turning computer power loose on the hard data of the Semantic Web." (Berners-Lee, 1999, S. 201 f.).

Auch Raymond (1999, S. 227) ist der Meinung, dass der *Open Source*-Ansatz in Zukunft Einfluß auf Gebiete jenseits der Software-Entwicklung haben wird: "I expect the open-source movement to have essentially won its point about software within three to five years. Once that is accomplished, and the results will be manifest for a while, they will become part of the background culture of non-programmers. At *that* point it will become more appropriate to try to leverage open-source insights in wider domains." In seiner Rede auf der XML Konferenz im Jahr 1999 hat Peter Murray-Rust die Definition der Semantik von XML als dringliches Desiderat angesprochen. Er warnt davor, dass im Zuge der Entwicklung diverser XML-Schemata ein „semantischer und ontologischer Krieg“ entstehen könnte, der nur durch unabhängige Gremien, die keinen finanziellen Interessen unterliegen, verhindert werden könne. Das Konzept der *Open Information* und eine sinnvolle Verknüpfung der verschiedenen Projekte in diesem Zusammenhang wäre unserer Meinung nach eine mögliche Lösung dieses unmittelbar bevorstehenden Problems.

Literatur

- BACA, MURTHA (Herausgeber) (1998): *Introduction to Metadata – Pathways to Digital Information*. Getty Information Institute.
- BERNERS-LEE, TIM (1999): *Weaving the Web – The Original Design and Ultimate Destiny of World Wide Web by Its Inventor*. San Francisco: Harper San Francisco.
- BRAY, TIM; PAOLI, JEAN UND SPERBERG-MCQUEEN, C. M. (1998): “Extensible Markup Language (XML) 1.0”. Technische Spezifikation, World Wide Web Consortium. Online verfügbar: <http://www.w3.org/TR/1998/REC-xml-19980210>.
- BRICKLEY, DAN UND GUHA, R.V. (1999): “Resource Description Framework (RDF) Schema Specification”. Technische Spezifikation, World Wide Web Consortium. Online verfügbar: <http://www.w3.org/TR/PR-rdf-schema/>.
- BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG, BUNDESMINISTERIUM FÜR WIRTSCHAFT UND TECHNOLOGIE (1999): “Innovation und Arbeitsplätze in der Informationsgesellschaft des 21. Jahrhunderts – Aktionsprogramm der Bundesregierung”. Online verfügbar: <http://www.iid.de/aktionen/aktionsprogramm/>.
- BUTHENUTH, ROGER UND MOCK, MARKUS U. (1992): “Abseits vom Kommerz – Die Philosophie des GNU-Projekts”. *c’t, Magazin für Computertechnik* (3): S. 62–65.
- DERTOUZOS, MICHAEL (1997): *What will be. How the new World of Information will change our Lives*. New York: HarperEdge.
- DIBONA, CHRIS; OCKMAN, SAM UND STONE, MARK (Herausgeber) (1999): *Open Sources: Voices from the Open Source Revolution*. Peking, Cambridge, Köln etc.: O’Reilly & Associates.
- ETTRICH, MATTHIAS (2000): “Wer kodiert? – Gedanken zur Freie-Software-Szene”. *iX, Magazin für professionelle Informationstechnik* (1): S. 112–115.
- FEUERBACH, HEINRICH T. UND SCHMITZ, PETER (1999): “Freiheitskämpfer – Entwicklung freier Software gegen Patentierung”. *c’t, Magazin für Computertechnik* (16): S. 79–81.
- FLUCKIGER, FRANÇOIS (1996): *Multimedia im Netz*. München, London, Mexiko etc.: Prentice Hall.
- HASEBROOK, JOACHIM (1995): *Multimedia-Psychologie*. Heidelberg, Berlin, Oxford: Spektrum.

- HUDGINS, JEAN; AGNEW, GRACE UND BROWN, ELIZABETH (1999): *Getting Mileage out of Metadata – Applications for the Library*, Band 5 von *LITA Guides*. Chicago: American Library Association.
- ISO10179 (1996): “Information Processing – Processing Languages – Document Style Semantics and Specification Language (DSSSL)”. Internationale Norm, International Organization for Standardization, Genf. Online verfügbar: <http://www.ornl.gov/sgml/wg8/>.
- ISO10744 (1997): “Information Processing – Hypermedia/Time-Based Structuring Language (HyTime) – Second Edition”. Internationale Norm, International Organization for Standardization, Genf. Online verfügbar: <http://www.ornl.gov/sgml/wg8/>.
- ISO8879 (1986): “Information Processing – Text and Office Information Systems – Standard Generalized Markup Language”. Internationale Norm, International Organization for Standardization, Genf.
- ISO/IEC13250 (1999): “Information Technology – Document Description and Processing Languages – Topic Maps”. Internationale Norm, International Organization for Standardization, Genf. Online verfügbar: <http://www.ornl.gov/sgml/wg4/>.
- LASSILA, ORA UND SWICK, RALPH R. (1999): “Resource Description Framework (RDF) Model and Syntax Specification”. Technische Spezifikation, World Wide Web Consortium. Online verfügbar: <http://www.w3.org/TR/REC-rdf-syntax/>.
- LOBIN, HENNING (2000): *Informationsmodellierung in XML und SGML*. Berlin, Heidelberg, New York etc.: Springer.
- MARCHIORI, MASSIMO (1998): “The Limits of Web Metadata, and beyond”. *Computer Networks and ISDN Systems* (30): S. 1–9. (auch: Proceedings of the 7th International World Wide Web Conference, Brisbane, Australien).
- NIELSEN, JAKOB (1996): *Multimedia, Hypertext und Internet. Grundlagen und Praxis des elektronischen Publizierens*. Braunschweig, Wiesbaden: Vieweg.
- O’REILLY, TIM (1999): “Hardware, Software, and Infoware”. In: DiBona et al. (1999), S. 189–196.
- PFAMMATTER, RENÉ (Herausgeber) (1998): *Multi Media Mania: Reflexionen zu Aspekten Neuer Medien*. Konstanz: UVK Medien.
- RAGGETT, DAVE; HORS, ARNAUD LE UND JACBOS, IAN (1997): “HTML 4.0 Specification”. Technische Spezifikation, World Wide Web Consortium. Online verfügbar: <http://www.w3.org/TR/REC-html40/>.

- RATH, HANS HOLGER (1999): "Mozart oder Kugel – Mit Topic Maps intelligente Informationsnetze aufbauen". *iX, Magazin für professionelle Informationstechnik* (12): S. 149–155.
- RAYMOND, ERIC S. (1999): *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Peking, Cambridge, Farnham etc.: O'Reilly & Associates.
- STALLMAN, RICHARD M. (1999): "The GNU Operating System and the Free Software Movement". In: DiBona et al. (1999), S. 53–70.
- STEINMETZ, RALF (1999): *Multimedia-Technologie – Grundlagen, Komponenten und Systeme*. New York, Berlin, Heidelberg etc.: Springer, 2. Auflage.
- VIXIE, PAUL (1999): "Software Engineering". In: DiBona et al. (1999), S. 91–100.
- VON HAAREN, KURT UND HENSCHKE, DETLEF (Herausgeber) (1997): *Arbeit im Multimedia-Zeitalter – Die Trends der Informationsgesellschaft*. Hamburg: VSA-Verlag.
- WEIBEL, S.; KUNZE, J.; LAGOZE, C. UND WOLF, M. (1999): "Dublin Core Metadata for Resource Discovery". Network Working Group, Request for Comments (RFC) 2413. Online verfügbar: <http://info.internet.isi.edu/1s/in-notes/rfc/files>.
- WEIZENBAUM, JOSEPH (1997): "Die Mythen der Informationsgesellschaft". In: von Haaren und Henschke (1997), S. 31–37.