

Elektronisches Publizieren in den achtziger Jahren
im Kontext technologischer Entwicklungslinien
unter Berücksichtigung sozialer Interessen

Freie wissenschaftliche Arbeit zur Erlangung des Grades
eines Magister Artium am Fachbereich Kommunikations-
wissenschaften der Freien Universität Berlin
am Arbeitsbereich Informationswissenschaft

eingereicht von Knud Böhle bei
Dr. Ralf-Dirk Hennings

Berlin, im Juli 1985

INHALTSVERZEICHNIS

1. EINLEITUNG	1
2. Annäherung an das Elektronische Publizieren	4
2.1 Der Bezug des Publizierens auf Öffentlichkeit	4
2.2 Publizieren als sozio-technische Figuration	6
2.2.1 Die Publikationskette	7
2.2.2 Ein Hypothesenraster	11
2.2.3 Standpunkte zum Elektronischen Publizieren	12
2.2.3.1 Der Standort des CAP	14
2.2.3.2 Verlegerstandpunkt	17
2.2.3.3 Standpunkt der Informationsindustrie	20
2.3 Elektronisches Publizieren im Rahmen der Textkommunikation	23
3. Vorüberlegungen zur SGML	29
3.1 Die Beziehung Autor - Verlag - Satzbetrieb	29
3.1.1 Das Rationalisierungspotential	29
3.1.2 Fremddatenübernahme	33
3.1.3 Die Autor - Verlagsbeziehung	40
3.1.4 Das Verhältnis Satz - Verlag	42
3.2 Ein allgemeines Modell der Dokumenterstellung	47
3.3 Entwicklungsschritte zur SGML	56
4. Elektronisches Publizieren mit der SGML	66
4.1 Die ISO/DIS 8879 (SGML)	66
4.1.1 Versprechungen der SGML	66
4.1.2 Zu Geschichte und Stand	69
4.1.3 Die Bauform der SGML	72
4.1.3.1 Kriterien einer standardisierten Markupsprache	72
4.1.3.2 Dokumenttypdefinition und SGML-Parser	75
4.1.3.3 Das Dokumentmodell der SGML	76
4.1.3.4 Markup Minimierung	83
4.1.3.5 SGML-Implementation und -Parser	84
4.1.3.6 Zwischenbetrachtung	85
4.2 Anwendungen der SGML	86
4.2.1 Dokumenttypen und Auszeichnungsrichtlinien	87
4.2.1.1 Das Electronic Manuscript Project	87
4.2.1.2 STRUKTEXT	89
4.2.1.3 DAPHNE	91
4.2.2 Anwendersoftware zur SGML	95
4.2.2.1 Markup mit einfachen Textverarbeitungssystemen	95
4.2.2.2 Konvertierprogramme	96
4.2.2.3 Intelligente Editoren	98
5. Schlußbetrachtungen	102
5.1 Überlegungen zur Akzeptanz der SGML	102
5.2 Übergreifende Ergebnisse und offene Fragen	109

6. Literaturverzeichnis	112
Anhang A. Rechercheauswertung	134

ABBILDUNGSVERZEICHNIS

Abbildung	1. Die Publikationskette	8
Abbildung	2. Informationsbanken zum Elektronischen Publizieren. . .	21
Abbildung	3. Ort des Elektronischen Publizierens I	25
Abbildung	4. Ort der Elektronischen Publizierens II	27
Abbildung	5. Stand der ISO/OSI Normung.	49
Abbildung	6. Schichtenmodell der Dokumenterstellung	51
Abbildung	7. Ablaufmodell der Dokumenterstellung	54
Abbildung	8. Logische Struktur des Dokuments "Arbeit"	77
Abbildung	9. Symbole der Graphendarstellung	93
Abbildung	10. Dokumenttyp "Brief"	94
Abbildung	11. Integriertes Elektronisches Publiziersystem	110
Abbildung	12. Rechercheauswertung I	137
Abbildung	13. Rechercheauswertung II	138

VORWORT

Die vorliegende Arbeit ist nicht nur eine schriftliche Arbeit zum Elektronischen Publizieren, sie ist von der Herstellungsseite auch ein Stück praktische Einübung in das Elektronischen Publizieren.

Teile der Arbeit sind zunächst auf dem einem IBM AT 02 erfaßt worden mit dem Textverarbeitungsprogramm *Euroscript*. In dem System habe ich darauf verzichtet, Formatierbefehle zu verwenden, d.h. einfachen Fließtext eingegeben.

Mit einem Filetransferprogramm wurden die Daten dann vom PC auf einen Großrechner, eine IBM 3081 überspielt. Dazu war es notwendig, den Textfile auf reinen Standard ASCII-Zeichensatz zu beschränken, d.h konkret die Umlaute zu codieren, da das Filetransferprogramm sonst "Blanks" an ihre Stelle setzt. Weiter konnte das Transferprogramm den Fließtext nicht so ohne weiteres übertragen, weil es mit festen Recordlängen arbeitet. Das Zeichen für Recordende (EOR) ist für das Transferprogramm Carriage Return/Linefeed. Derselbe Code wird in *Euroscript* für Absatzende verwendet. Bevor ein Absatzende im Text erreicht wurde, war in der Regel die maximale Länge eines Records, die das Transferprogramm akzeptiert, überschritten. Ich habe dann von anderen Nutzern dieses Transferprogramms ein kleines Basicprogramm auf Diskette erhalten, daß den Fließtext in Records von 80 Zeichen "zerhackt". Da das Programm für einen anderen Zweck geschrieben war, bedurfte es noch kleiner Anpassungen (ca.4 Stunden) am Programm, die mit fremder Hilfe vorgenommen werden konnten. Als das Programm dann fehlerlos lief, wurde offenkundig, daß Dateien von mehr als 7000 Zeichen nicht auf dem Großrechner angenommen wurden. Der Fehler ist, soweit ich weiß noch nicht behoben. Die Folge war, daß längere Dateien in kleinere vor dem Transfer geteilt werden mußten.

Da beabsichtigt war, die Texte sowieso mit der DCFGML (Document Composition Facility Generalized Markup Language) weiterzuverarbeiten, habe ich mich dann von der Transfermethode abgewandt und von einem Terminal 3179G aus direkt am Großrechner weitererfaßt. Die bessere Qualität des Terminalmonitors hat mir den Abschied von der Textverarbeitung am PC, trotz seiner Mobilität, erleichtert.

In dem Editor des Systems gibt man - mit dem üblichen Komfort solcher Editoren - seinen Text ein. In Hinblick auf die Verarbeitung mit der DCFGML allerdings sind noch logische Auszeichnungen in den Text einzustreuen.

Diese logischen Auszeichnungen werden von dem Formatierer SCRIPT/VS verarbeitet und für die gestaltete Ausgabe auf dem Bildschirm oder

auf angeschlossenen Druckern aufbereitet. Die vorliegende Ausgabe geht auf den Laserdrucker 6670 (Modell II) von IBM zurück.

Für die Graphiken kamen noch zwei weitere Verfahren in Anwendung. Durch entsprechende Angaben auf Ebene der DCFGML wurde im Ausdruck Platz für Graphiken freigeschlagen. Diese wurden auf einer XEROX 8012 Workstation, einem Bürosystem, das Text und Grafik integriert, nach meinen Vorlagen erstellt. Eine Abbildung, einem Buch entnommen, wurde kopiert und eingefügt. Die nicht ganzseitigen Abbildungen wurden in den Text an den entsprechenden Stellen eingeklebt. Das System kann - vor allem unter Einschluß des PCs - leicht ironisierend als Hybridsystem angesehen werden. Andererseits liegt die Arbeit nun in einer m.E. ganz ansprechenden äußeren Form vor *und* als maschinenlesbarer Text, der ohne großen Aufwand aktualisiert, verändert und transferiert werden kann, da der logisch ausgezeichnete, nicht formatierte Text einen reinen ASCII-File darstellt.

1. EINLEITUNG

Seit Anfang der achtziger Jahre ist Elektronisches Publizieren ein Thema mehr, das im "revolutionären" Kontext der Telematik, d.h. des Zusammenwachsens der Telekommunikation und der Informatik aufgekomen ist.

Anders als etwa bei Videotex ist noch in der Diskussion, was mit Elektronischem Publizieren eigentlich auf den Begriff gebracht werden soll. Es gibt deshalb kaum einen Artikel zum Elektronischen Publizieren, der ohne eine zumindest kurze Vorverständigung auskäme.

Elektronisches Publizieren in den achtziger Jahren bezeichnet ein vorschießendes Konzept, dem nur partiell eine Realität entspricht. Die technologischen Entwicklungslinien der Telematik bilden den Kontext, die übergreifende Entwicklungsdynamik. Soziale Interessen erst gestalten den Technikeinsatz im sich damit verändernden sozialen Raum.

Das Thema, das zunächst vielleicht überfrachtet erscheint durch seinen Bezug auf technologische Entwicklungslinien und soziale Interessen, qualifiziert sich bei genauerem Hinsehen gerade dadurch als informationswissenschaftliches Thema, da Informationswissenschaft genau diesen Zusammenhang von Technik und Gesellschaft aufzuklären zu ihrem zentralen Gegenstand hat.

"Was nottut, ist die Schaffung von Bewußtsein und Kapazitäten, um diese umgreifende und durchdringende Bewegung (der Informatisierung - K.B.) zu begreifen. Dazu gehört

- ein kontinuierliches *Monitoring* der Entwicklung selber, etwa in Form der laufenden Beobachtung, Verfolgung und Auswertung (...)
- eine kontinuierliche *Folgenabschätzung*
- eine kontinuierliche *Strategiefortschreibung*.

Ansätze dafür gibt es einige, Informationswissenschaft gehört zweifelsohne ebenso dazu wie die Kreise der Begleitforschung zu den "Neuen Medien",... " /Wersig 1984, 85/.

Wenn man einen Untersuchungsbereich Elektronisches Publizieren abgrenzen will, um der Uferlosigkeit der Fragestellungen entgegenzutreten, steht man unwillkürlich vor dem Dilemma, Bereiche und Aspekte auszuklammern, die durch das integrative Potential der Technologie angekoppelt werden, eigentlich auch dazu gehören.

Es ist nicht nur die "schlechte" Unendlichkeit der Zusammenhänge, die zur Schwerpunktsetzung zwingt, es ist auch der Charakter des Elektronischen Publizierens als sozial *und* technisch bestimmter Handlungsraum, der im Rahmen einer solchen Arbeit eine Schwerpunktsetzung notwendig macht, wenn man über Allgemeinplätze hinauskommen will.

Ich habe mich deshalb für ein exemplarisches Vorgehen entschieden. Im Zentrum dieser Arbeit steht eine ISO Norm, die ISO/DIS 8879 - Standard Generalized Markup Language -. Diese Entscheidung begründe ich damit, daß die Standard Generalized Markup Language in mehrerer Hinsicht einen Schlüssel zum Elektronischen Publizieren darstellt. Sich mit der Standardisierung einer Computersprache für die Dokumentverarbeitung zu befassen, macht einmal Sinn, weil die Durchsetzung der Telematik - und in diesem Kontext das Elektronische Publizieren - ganz entscheidend an die Entwicklung von Standards gebunden ist. Die Rolle des ISO/OSI-Modells kann in diesem Zusammenhang m.E. kaum überschätzt werden.

Sich mit einem Standard in diesem Bereich zu beschäftigen macht auch deshalb Sinn, weil die Normen, bildlich gesprochen, sozio-technische Kristallisationspunkte darstellen. Das fängt bei der Interessenkonstellation an, die aus einem Normvorhaben eine Norm werden läßt. Das geht weiter mit der Entwicklung von Hard- und Software-systemen, die sich auf die Norm beziehen und hat eine weitere Dimension der Akzeptanz bei den Anwendern und schließlich den End-nutzern.

Die ISO/DIS 8879 ist weiter eine Schlüsselentwicklung für den Bereich des Elektronischen Publizierens, weil ihr Anspruch, wieder bildlich gesprochen, der ist, das Schmiermittel des Elektronischen Publizierens zu sein.

Sie spielt für die Erstellung und Weiterverarbeitbarkeit Elektronischer Manuskripte ebenso eine Rolle wie für den Aufbau einer Datenbasis, aus der elektronische Publikationen gewonnen werden können, egal ob für zentrale Lieferformen (Host) oder dezentrale Formen wie CD-ROM. Sie ist somit auch ein Bindeglied zwischen dem produktionsorientierten "Computer Aided Publishing" und dem lieferformorientierten "Document Delivery", durchzieht m.a.W. die ganze Publikationskette.

Die Arbeit ist technikorientiert. Sie stellt den Versuch eines Nicht-Informatikers dar, sich soweit mit einem Technikausschnitt vertraut zu machen, daß Übervereinfachungen vermieden und Rückschlüsse auf die soziale Bedeutung des Technikausschnitts möglich werden.

Die Arbeit teilt sich in vier große Blöcke. In einem ersten wird eine Annäherung an den Gegenstand "Elektronisches Publizieren" versucht. Dabei kommt es mir darauf an, Elektronisches Publizieren nicht technisch zu definieren, sondern einmal funktional bezogen auf die Kategorie der Öffentlichkeit und zum anderen "strukturell" bezogen auf "Elektronisches Publizieren" als sozio-technische Figuration, d.h. einen bestimmten Handlungszusammenhang. Diesen schließe ich einmal von dem Konzept der "Publikationskette" her auf, indem ich die Stationen einer Publikation durchgehe. Zum anderen präpariere ich drei standortgebundene Sichtweisen auf das Elektronische Publizieren heraus (Druckbereich, Verlagsbereich, Informationsindustrie) und stelle in dem Zusammenhang einschlägige Definitionen zum Elektronischen Publizieren vor. Ich schlage danach vor, Elektronisches Publizieren in den Rahmen der Textkommunikation zu stellen.

In dem nächsten Kapitel sammle ich Grundlagen für die Beschäftigung mit der SGML. Dazu rechne ich einmal eine genauere Betrachtung der Autor - Verlag - Satz - Beziehung. Zu den Grundlagen gehört ganz wesentlich noch die Entwicklung eines allgemeinen Modells der Dokumenterstellung, das am ISO-Referenzmodell ansetzt und den Dokumenterstellungprozeß als eine Differenzierung der Anwendungsschicht modelliert. Als dritte Vorüberlegung wird die Forderung einer "Standardisierten Verallgemeinerten Auszeichnungssprache" aus den Defiziten üblicher Textverarbeitungssoftware rekonstruiert.

Im dritten Block wird die SGML ausführlich dargestellt. Nach der inhaltlichen Klärung, die mit Beispielen erleichtert werden soll, wird der Stand der Anwendungen ausgebreitet. Darunter fallen Projekte zur Entwicklung von Autorenrichtlinien ebenso wie Softwareentwicklungen auf Basis der SGML.

In dem Schlußkapitel werden die Ergebnisse der Arbeit unter dem Gesichtspunkt der Akzeptanzbedingungen der Norm zusammengetragen und kurz angeschnitten wie eine angewandte Informationswissenschaft an diesen Ergebnissen ansetzen könnte.

Die Literaturliste habe ich versucht nicht ausufern zu lassen, was seit der "Online-Revolution" nicht schwer fiel. Als "Kompensation" und als Überblick über die Publikationstätigkeit zum Elektronischen Publizieren habe ich die Ergebnisse von elf Datenbankrecherchen zum Thema leicht bearbeitet und interpretiert angefügt. Zumindest Themen, um die die Diskussion zum Elektronischen Publizieren kreist, und Hinweise auf Datenbanken, in denen man fündig wird, lassen sich daraus entnehmen.

2. ANNÄHERUNG AN DAS ELEKTRONISCHE PUBLIZIEREN

In diesem Kapitel soll die Annäherung an das Elektronische Publizieren darin bestehen, daß zunächst der Bezug des Elektronischen Publizierens zum Publizieren allgemein deutlich gemacht wird, so dann wird Publizieren als eine sozio-technische Konfiguration herausgearbeitet. Ausgehend von bestimmten Standorten in dieser Figuration schließen sich Definitionen des Elektronischen Publizierens auf. Dieses Kapitel abschließend wird Elektronisches Publizieren in den Rahmen der Textkommunikation gestellt, um den Gegenstand für den weiteren Verlauf der Arbeit operational, gerade für die Erörterung auch technischer Fragen geeignet, einzuschränken.

2.1 DER BEZUG DES PUBLIZIERENS AUF ÖFFENTLICHKEIT

"Elektronisches Publizieren", das eingedeutschte Äquivalent zum "electronic publishing", setzt sich automatisch in ein Verhältnis zum nicht-elektronischen Publizieren. Publizieren (publishing) wird von dem Office of Arts and Libraries (OAL)¹ definiert als

"... the systematic open communication of information and ideas through recognised channels in such a way as to permit retention and archiving for future use."/Impact electronic publishing 1983, 281/

Diese Definition hat für sich, daß sie Publizieren auf eine Struktur bzw. Funktion bezieht und nicht auf die zur Erbringung der Funktion eingesetzten Techniken.

Diese Definition vermeidet damit ebenfalls die Rede von traditionellen und "neuen" Medien, die sofort zu begrifflichen Schwierigkeiten führt. Wenn ein "altes" Medium mit einer neuartigen Technologie erstellt wird (z.B. Faksimilezeitung, Elektronische Zeit-

¹ Von dem Library and Information Services Council (LISC) wurde die Studie angeregt, die von einer kleinen Arbeitsgruppe innerhalb der OAL durchgeführt wurde und dem LISC und dem Minister of Arts (GB) zu berichten hatte /vgl. Impact electronic publishing 1983, 283/.

schrift), spaltet sich dann das Medium in "alt" und "neu"? Auch "Medienmix"² geht nicht in der Gegenüberstellung "alt" "neu" auf. Das entscheidende Argument, den Ausdruck "neue" Medien nicht für einen tauglichen Begriff zu halten, sehe ich in der Entwicklung von Netzen, die das Konzept des Mediums, *eines Dritten*, "unterlaufen", sobald sie *direkte* Kommunikation, Interaktivität, gestatten.³

Die systematisch offene Kommunikation über anerkannte Kanäle, Teil der obigen Definition, ist eine Formulierung für den notwendigen Bezug des Publizierens auf Öffentlichkeit.⁴ Damit wird die Betrachtung für Fragen nach dem "Strukturwandel der Öffentlichkeit" geöffnet oder anders die Frage, wie spezifische Kommunikationsmittel zur Herstellung von Öffentlichkeit und ihrer Funktionsfähigkeit beitragen, gestellt.⁵

Nicht unwichtig ist auch der letzte Teil der Definition, der vom Publizieren erst dann spricht, wenn dafür gesorgt ist, daß eine

² Die OAL Studie weist auf eine "interessante" Mischung hin: "... the BBC, using videodisc storage, has recently produced a guide to British birds, which uses film, text and sound recordings. It shows birds stationary, in flight and on the nest. Through its twin channels of stereo sound, it offers both a commentary and recording of bird noises and sounds. A teletext track provides titles for the pictures and additional text for descriptive titles, while an index on the disc indicates which teletext page numbers contain additional information" /ebd., 284/.

³ Vgl. zur Problematik des Medienbegriffs /Wersig 1985 b, 14-20, bes.17/

⁴ Die Kategorie Öffentlichkeit "gehört spezifisch zur 'bürgerlichen Gesellschaft', die sich zur gleichen Zeit als Bereich des Warenverkehrs und der gesellschaftlichen Arbeit nach eigenen Gesetzen etabliert" /Habermas 1978, 15/.

⁵ Unter dem Stichwort "Kabeldemokratie" haben Vowe und Wersig die Position vertreten, daß die neuen Informations- und Kommunikationstechniken als Chance verstanden werden müssen, die Funktionsfähigkeit des gesellschaftlichen Kommunikationssystems zu verbessern. "Die Telematik stellt die sachtechnische Option dar, um das gesellschaftliche Kommunikationssystem den neu sich entfaltenden Anforderungen anzupassen ... Notwendig ist die Konzeption von gesellschaftsorganisatorischen und im Sinne der Defizitabdeckung 'richtigen' Gestaltungs- und Einsatzformen dieser sachtechnischen Option" /Vowe, Wersig 1983, 276f/.

Information so vorgehalten wird, daß auf sie zeitpunktunabhängig zugegriffen werden kann. Damit erst wird dem Kriterium der allgemeinen Zugänglichkeit, mit dem Öffentlichkeit steht und fällt, genügt.

Dem Publizieren, wie es in der Definition gefaßt wurde, liegt eine sozio-technische Figuration⁶ zugrunde.

An dieser Figuration wirken die Produzenten von Informationen mit, die an den technischen und organisatorischen Produktions- und Distributionsprozessen Beteiligten und die Rezipienten (das Publikum), die Informationen (möglicherweise) wieder für die Produktion von Informationen oder in anderen Handlungskontexten - politischen, wissenschaftlichen, rechtlichen etc. - einsetzen.

Von daher ist es einsichtig, daß der Publikationsprozeß gerne als Kette ("chain" oder "circle" respektive) ins Bild gebracht wird.⁷ Elektronisches Publizieren bezeichnet den Übergang von der traditionellen zur elektronischen Kette oder, etwas schärfer formuliert, einen (auch) technikinduzierten Wandlungsprozeß im Muster öffentlicher Kommunikation und ihrer materiellen Voraussetzungen.

2.2 PUBLIZIEREN ALS SOZIO-TECHNISCHE FIGURATION

In diesem Abschnitt wird die sozio-technische Figuration des Publizierens überblickartig dargestellt mit der Absicht, Handlungsträger und Faktoren zu identifizieren, die den Umbruch zum Elektronischen Publizieren befördern oder auch hemmen.

⁶ Norbert Elias führt diesen Begriff ein, um ein einfaches begriffliches Werkzeug zu schaffen für sich wandelnde Muster, die Akteure in ihren Beziehungen zueinander bilden. /Elias 1971, 139 -145/

⁷ Als Studien, die sich mit der Publikationskette befassen, sind herauszuheben die bereits zitierte /Impact electronic publishing 1983/, /Oakeshott 1983/ und /Oakeshott, White 1984/.

2.2.1 Die Publikationskette

Auf der folgenden Seite habe ich versucht, die Publikationskette graphisch darzustellen.

Die Stationen, die eine Veröffentlichung durchläuft (vom Autor und seinem Manuskript zur Publikation und von der Publikation bis zum Nutzer), sind als Rechtecke dargestellt, die Kommunikations- und Transportwege als Pfeile. Dort, wo elektronische Informations- und Kommunikationstechniken ins Spiel kommen (können), ist das durch eine Schraffierung hervorgehoben.

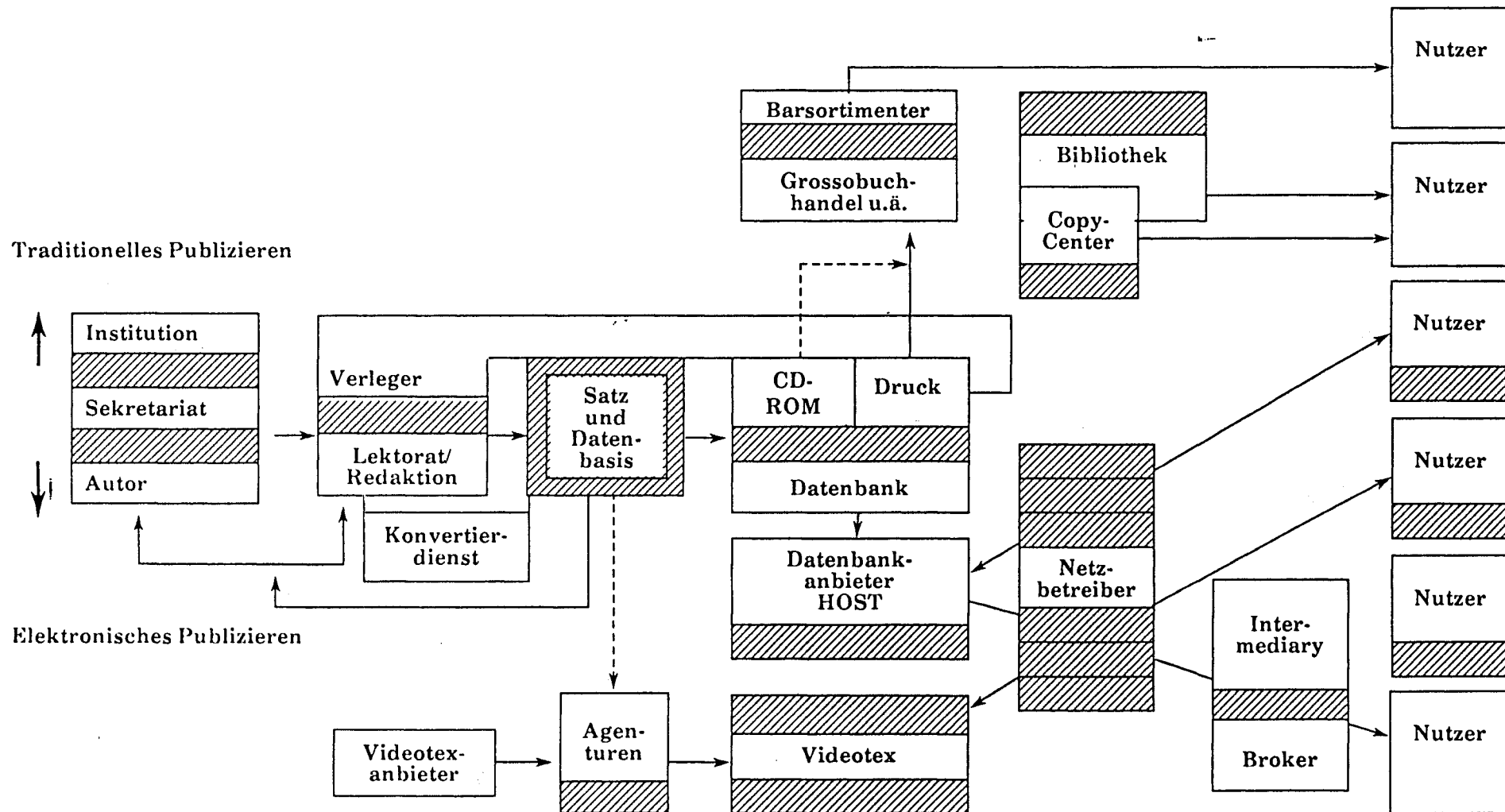
Der konventionelle Weg einer Veröffentlichung vom Manuskript des Autors bis zur Publikation beim Leser bzw. Nutzer beginnt beim Autor, der die Rohform einer Publikation, sein Manuskript erstellt. Bezogen auf die Vermarktung dieses Produktes und seine Veröffentlichung tritt er/sie in eine kommunikative Beziehung mit einem Verlag bzw. dessen Lektorat. In dieser Kommunikation werden inhaltliche und stilistische Veränderungen am Manuskript (von rechtlichen Aspekten einmal abgesehen) vereinbart und vorgenommen.

Wenn der "Input" in fertiger Form vorliegt, beginnt die eigentliche Herstellung des Druckerzeugnisses. Dazu vergibt der Verlag Aufträge an Druckbetriebe. Nicht selten besitzt ein Verlag eine Druckerei oder andersherum. Konventionell wird der Manuskripttext in einem Satzbetrieb neu erfaßt und gesetzt, d.h. nach den Regeln der "schwarzen Kunst" und den Richtlinien der Auftraggeber für die Ausgabe gestaltet. Als Zwischenprodukt werden die Satzfähnen erzeugt, die dann zur Korrektur an Autor und Verlag gehen (siehe entsprechenden Pfeil in der Abbildung). Diese Schleife wird beendet, wenn der Autor sein *Imprimatur* gibt, d.h. seine Erlaubnis zum Druck erteilt. Danach wird die Druckvorlage vervielfältigt, es entsteht das Druckerzeugnis in einer bestimmten Auflagenhöhe. Damit ist der Produktionsprozeß abgeschlossen.

Nun beginnt die dritte Phase des Vertriebs oder der Distribution. Der Buchhandel, selbst noch eine zu differenzierende Funktionsgruppe, übernimmt diese Aufgabe. Sammelstellen, wie Bibliotheken, und Leser oder Endnutzer erwerben die Veröffentlichungen über den Buchhandel.

In diesem Schema tritt die Informations- und Kommunikationstechnik praktisch als Satzrechner in Erscheinung, wenn man nicht bis zum Bleisatz zurückgehen will. Alle anderen Techniken sind noch mechanisch und auf den Datenträger Papier orientiert.

Selbstverständlich ist die gerade beschriebene "Idylle" längst nicht mehr die ganze Geschichte. Das fängt damit an, daß die Kopierer praktisch elektronisch gesteuerte "non-impact"-Drucker sind, die aus dem (wissenschaftlichen) Publikationsbereich zur Zeit kaum



< INPUT →

> < DATENBASIS > < Präsentationsformen →

> < Weiterverarbeitung >

< Produktion →

> < Distribution →

> < Nutzung → >

Shaded box = möglicher Einsatz neuer IKT

..... = mögliche Verbindungen

wegzudenken sind. Sie stehen in den Bibliotheken oder in Spezialbetrieben, den sogenannten Copy-Center(n).

Eine andere Entwicklung ist ebenfalls nicht mehr neu. Schon lange gibt es "Sekundärpublikationen" wie Referatedienste, die Informationen über Informationen bieten, um die Orientierung in der stetig ansteigenden Literaturmenge zu erleichtern.

Die Geschichte des Information Retrieval beginnt damit, daß diese Referatedienste auf den "Satzrechnern" in maschinenlesbarer Form vorlagen und man Techniken entwickelte, um die Registererstellung, die ja den Wert solcher Werke ausmacht, zu automatisieren.⁸ Wenn man es ganz einfach sagen will, ist das Information Retrieval mit der Idee geboren worden, den Verweismechanismus in beide Richtungen zu nutzen, d.h auch zum Auffinden der indexierten Stelle im Text. Mit dem öffentlichen Zugang zu diesen Datenbanken und der Fortentwicklung von Information Retrieval Software ist so nebenbei eine Nebenform des Elektronischen Publizierens entstanden.

"... we may define electronic publishing as using electronic means for the storage and display of textual material rather than audio or visual as in the case of radio and tv. As such, electronic publishing is not particularly new. It has been with us since the 1960's with the various indexing and abstracting services in various scientific disciplines." /Stern 1985, 2/

Ob Sekundärpublikationen dem Begriff Publikationen entsprechen oder nicht, ist in diesem Zusammenhang nicht so interessant wie das Prinzip, das hier zum Tragen kommt. Informationen / Publikationen werden elektronisch gespeichert und öffentlich zugänglich gemacht. Unter anderem Gesichtspunkt sind diese Datenbanken der erste Fall der Mehrfachnutzung eines elektronischen Datenbestandes. In der Abbildung führt dieser Strang vom Secondary Publishing über Satz/Datenbasis und Datenbank zum Datenbankanbieter.

Derselbe Weg führt aber auch vom Verlag zu Satz/Datenbasis und weiter zu den Ausgabeformen Druck, Datenbank, CD-ROM. Eine Datenübergabe für Videotex-Dienste ist eine weitere Option.⁹ Während

⁸ /Salton, McGill 1983/ gehen in dem zweiten Kapitel ihres Buches ausführlich auf Information Retrieval Systeme, die mit invertierten Dateien arbeiten, ein.

⁹ /Vgl. Sutter, Teichmann 1982/

Druckerzeugnis und digitale Bildplatte¹⁰ den üblichen Vertriebsweg gehen (können im Falle der Bildplatte), geht der Weg von der Datenbank wieder zum Host. Im Unterschied allerdings zu den Sekundärpublikationen geht es bei dem gerade beschriebenen Weg um Volltexte. Auch hier können die verschiedenen Ausgabe- bzw. Lieferformen parallel auftreten, d.h. dem Prinzip der Mehrfachnutzung einer Datenbasis gehorchen.

Aus der Abbildung wird auch deutlich, daß der Zugriff auf Datenbanken bei einem Host Kommunikationsnetze und eine technische Ausstattung bei den Nutzern oder zwischengeschalteten Professionen voraussetzt.

Ist die Mehrfachverwertung einer Datenbasis ein zentraler Ausgangspunkt des Elektronischen Publizierens, so ist das Elektronische Manuskript ein weiterer, nicht weniger wichtiger. Immer mehr Manuskripte werden elektronisch erstellt. Dabei spielen Arbeitsplatzrechner eine Rolle, Bürosysteme, die oft schon sehr weit fortgeschrittene Text-Graphik Integration anbieten, und Textverarbeitungssysteme auf Großrechnern. Damit ist im optimalen Fall erreicht, daß die elektronische Publikationskette geschlossen werden kann, vom Autor bis zum Nutzer.¹¹

Dieses Ablaufmodell ist schematisch und sollte praktisch nur die Hauptakteure der Figuration vorstellen. Es läßt sich bereits jetzt behaupten, daß sich mit dem Elektronischen Publizieren die Zahl der Akteure in der Figuration erhöht: Die Hersteller von Hard- und Software, seien sie auf die Bedürfnisse von Autoren, Verlegern, Druckbetrieben, Hosts, Bibliotheken oder Nutzern spezialisiert, kommen hinzu; Datenbank- und Videotex-Anbieter treten als Informationsanbieter neben die Verlage. Kommunikationsnetze erst machen die Nutzung der Dienste möglich. Publizieren wird Teil des informations-industriellen Komplexes.

¹⁰ Da auf die Bildplattentechnologie und ihren Einsatz nicht weiter eingegangen wird, hier nur einige Literaturhinweise: /Information Systems Consultants 1985/, /Hensel 1984/, /Hennings 1985/, /Hendley 1985/.

¹¹ In diese elektronische Kette können sich auch der Buchhandel und die Bibliotheken eingliedern. Ein Buchhändler nutzt eine Datenbank des Grossobuchhandels für seine Bestellaufträge und verkauft "CD-ROM", wie bereits jetzt "sprechende Bücher". Die Bibliothek übernimmt Informationsvermittlungstätigkeiten oder beteiligt sich an dem Abruf von Volltexten aus Datenbanken für ihre Nutzer, wie sie jetzt z.B. im Fernleihwesen Kopien zieht.

"Mit dem Schlagwort des informations-industriellen Komplexes wird versucht anzudeuten, daß für eine bestimmte Kommunikationslandschaft die Produzenten der informations- und kommunikationstechnischen Geräte, die Anbieter von Basis-Diensten, die Betreiber der Übermittlungswege und die Produzenten der flächendeckenden Dienste zu einem Komplex zusammenrücken, der zwar in sich strukturiert ist, in dem aber die einzelnen Komponenten derart voneinander abhängen, daß sie zwangsläufig zusammenwirken müssen." /Wersig 1985 a, 265/

2.2.2 Ein Hypothesenraster

Die Publikationskette stellte lediglich die "Spieler" vor, sagt aber nichts über die technischen und nichts über die sozialen Beziehungen der Akteure aus, die aus den "Spielern" erst eine Figuration machen. Ich möchte dazu im folgenden ein grobes Hypothesenraster aufstellen, das zwar einerseits aus einer Reihe recht bekannter Grundannahmen besteht, andererseits aber doch ausreicht, um die sozio-technische Konfiguration nicht zu übervereinfachen.

1. Der sozio-technische Wandel in der Konfiguration ist von der technischen Seite her im Kontext der Telematik zu sehen, deren allgemeines Entwicklungspotential Wersig in fünf Trends aufteilt:
 - "...eine stärkere Verschiebung von den transportgebundenen Kommunikationstechnologien zu den auf energetische Übertragung angelegten
 - eine stärkere Vereinheitlichung unterschiedlicher Darstellungsformen mittels durchgängiger Digitalisierung
 - eine stärkere Automatisierung gegenüber früheren Stadien der Mechanisierung (mit allen Formen der neuen Formen "maschineller Intelligenz")
 - umfangreichere Möglichkeit, Mitteilungen unterschiedlichster Form wahlweise gezielt oder verteilt zu übermitteln
 - stärkere Zeitpunktunabhängigkeit von einzelnen Stationen des Kommunikationsprozesses." /Wersig 1984, 38/
2. Der Einsatz der neuen Informations- und Kommunikationstechniken im Publikationsbereich ist in hohem Maße an infrastrukturelle Voraussetzungen gebunden, deren wichtigste sind:
Verbreitung von Mikrocomputern,

Ausbau der Kommunikationsnetze und -dienste,
Standardisierungen von Schnittstellen und Anwendungen,
Vermittlung des nötigen technischen Wissens in
Aus- und Weiterbildung ("computer literacy").

3. Die Akzeptanz und Durchsetzung der neuen Informations- und Kommunikationstechniken in einem bestimmten Anwendungsfeld wird davon abhängen,
daß die technischen Probleme, die mit der Integration und Anpassung der Basistechnologien an die Erfordernisse bestimmter Einsatzfelder entstehen, gelöst werden. Darunter fällt

- daß komfortable Hard- und Software die Technik im spezifischen Anwendungsbereich leicht handhabbar macht
- daß die Qualität der erzeugten neuen Produkte oder Dienstleistungen nicht hinter die Qualitätsstandards der gewohnten Angebote zurückfällt (bzw. spezielle Vorteile andere Nachteile "überkompensieren")

daß der Einsatz der neuen Techniken eine vergleichsweise bessere Bewältigung realer Probleme erwarten läßt;
daß sich der Einsatz neuer Technologie rechnet.

4. Der Einsatz der neuen Informations- und Kommunikationstechniken muß durch das Nadelöhr (nicht nur materieller) sozialer Interessen und sozialen Selbstverständnisses. Dabei lassen sich unterscheiden:

Interessen an Gebrauchswerten und Qualitäten, besonders bezogen auf den Umgang mit der Technik und die Darstellung der Informationen;

Interesse an der Funktionsfähigkeit der Öffentlichkeit bezogen auf das Informationsproblem (Bewältigung der Informationsmenge etc.) und auf das Postulat des allgemeinen Zugangs;

ökonomische Interessen: Rationalisierung der Produktion, Auslastung der Maschinerie, Erschließen neuer Märkte;

Interesse (nicht nur) "konservativer" Akteure im Publikationsbereich am status quo .

2.2.3 Standpunkte zum Elektronischen Publizieren

Dies Hypothesenraster soll nun dadurch gefüllt werden, daß reichsspezifische, standortgebundene Sichten des Elektronischen Publizierens vorgeführt werden.

Es ist interessant zu sehen, daß in der Literatur drei Perspektiven auf das Elektronische Publizieren vertreten sind. Die erste ist produktionsorientiert und häufig an das Konzept des "Computer Aided Publishing" (CAP) oder "Computer Assisted Publishing" gebunden. Sie ist wesentlich auf Fortschritte in der Druckindustrie, besonders im Satzbereich, die sich aus dem Einsatz der Mikroelektronik ergeben, gerichtet.

Der zweite Standpunkt, der des Verlegers, setzt das CAP voraus, geht schon einen Schritt weiter, indem er, zwar gedanklich noch auf Printmedien bezogen, die Idee elektronischer Lieferformen entwickelt.

Die dritte Position, die der Informationsindustrie, kennt keine Orientierung auf die Printmedien mehr und denkt in Termini elektronischer Dienstleistungen.

Joan Smith folgt einer ähnlichen Einteilung nur von einer Sachlogik aus gesehen, die den Standorten korrespondiert.

"Computer-assisted publishing is where the computer is used for any of the publication stages to result in a typeset document (hard copy); electronic publishing is where the document remains in an electronic environment and is published on a screen (soft copy). With database publishing the elements of a document may be retrieved and published in hard or soft copy, either alone, as a complete document, or in combination with other elements" /Smith 1986e, 195/.

Manfred Krüger unterscheidet nach sachlichen Gesichtspunkten zwei Definitionen von Elektronischem Publizieren: eine produktionsorientierte und eine lieferformorientierte. Anders als Joan Smith, die praktisch innerhalb der Lieferformen zwischen elektronischer Dokumentauslieferung (bei ihr electronic publishing) und dem selektierenden Zugriff auf Information Storage & Retrievalsysteme unterscheidet, teilt Krüger die Lieferformen in Datenbanken und elektronische Datenträger ein.

"Elektronisches Publizieren - Definition 1
(produktionsorientiert)

Elektronische Texterfassung und Umsetzung, d.h. Substitution der beiden Teilprozesse für 'herkömmliche Manuskripterstellung' und 'herkömmliche Satzerstellung' durch die Integration zu einem einheitlichen, elektronisch bestimmten Prozeß, ggf. erweitert durch elektronische Drucksysteme anstelle konventioneller Druckverfahren" /Krüger 1986b, 27/.

"Elektronisches Publizieren - Definition 2
(lieferformorientiert)

von elektronisch gespeicherten Texten (Dokumenten) durch elektronisch zugängliche Datenbanken, die im Unterschied zur konventionellen Vervielfältigung und Verbreitung von Printmedien neuartige Leistungen anbieten (Retrieval, Auszüge aus Dokumenten usw.) und damit partiell mit den Printmedien in eine bisher unbekannte Konkurrenz treten. Daneben treten Angebote elektronischer Datenträger" /Krüger 1986b, 28/.

Über die Richtigkeit, die Produktionsseite von der Angebotsseite zu trennen, gibt es offensichtlich Übereinstimmung. Was meiner Meinung nach in den Definitionen sowohl von Smith als auch von Krüger enthalten ist, ist die Unterscheidung von "innovativen" Formen des Elektronischen Publizierens und Formen, die ein Printprodukt lediglich elektronisch verfügbar machen, ohne aber im Grunde seine Leistungsmerkmale zu erweitern.

Wenn man standortbezogene Sichten des Elektronischen Publizierens aufsucht, werden die *drei* Perspektiven auf das Elektronische Publizieren deutlich.

2.2.3.1 Der Standort des CAP

Es lassen sich drei verschieden weite Kreise um das Elektronische Publizieren ziehen. Der engste Kreis wird beschrieben, wenn nur der Produktionsprozeß von Publikationen in den Blick rückt.

Der Einsatz der neuen Informations- und Kommunikationstechniken im Druckbereich gibt den Standort an, von dem aus Elektronisches Publizieren produktionsorientiert in den Blick rückt. Hier stehen innerbetriebliche wie zwischenbetriebliche Rationalisierungsschübe im Vordergrund, die den gesamten Bereich auf eine neue technologische Basis stellen und auch die Schnittstelle zu den Kunden neu definieren.¹²

Das gewachsene Potential, für den Markt elektronischer Medien und Dienstleistungen zu produzieren und so den Druckbereich an die Informationsindustrie anzubinden, kann als nicht-intendierte Folge angesehen werden, die erst allmählich als Chance erkannt wurde. In einer Studie im Auftrag des Bundesverbands Druck wird der Anschluß der Druckbetriebe an die Informationsindustrie propagiert:

¹² Vgl. zu den jüngsten Entwicklungen im Druckbereich beispielsweise: Deutscher Drucker 1986, Nr.13: Information und Fortschritt weltweit. Drupa 86 Technologie-Dokumentation oder die vergleichbaren Sonderausgaben zur DRUPA 86 des Polygraph oder des Druckspiegels. Zu den Rationalisierungskonflikten vgl. statt anderer /Weber, 1982/.

"Nahezu alle Bereiche der Druckindustrie werden von den Entwicklungen auf den Gebieten Elektronik und Datenverarbeitung tangiert. Die größten Veränderungen gehen jedoch z.Z. im Bereich der Satzherstellung vor sich. Satzbetriebe sollten das vorhandene Wissen um Gestaltung und die in ihrem Betrieb verfügbare digital gespeicherte Informationsmenge als günstige Voraussetzung erkennen, um mit einem frühen Einstieg in die neuen Medien die neu zu verteilenden Marktpositionen zu besetzen. Eine weitere Entwicklung sollte aus der vorhandenen Ausstattung eines Satzbetriebes mit digitaler Datenspeicherung resultieren. Die Nutzung dieser Rechnerkapazität, auch in bezug auf Datensammlung und -schöpfung, deren Manipulation und Neuverwertung, birgt viele völlig neue Möglichkeiten zur Erweiterung der Betriebsaktivitäten, die allerdings auch fachspezifisches Wissen (Datenverarbeitung) erfordern." /Teichmann u.a. 1982, 33/

Die Chance wird aber auch als Notwendigkeit gesehen:

"Die heutige Druckindustrie wird sich zukünftig zu einer *Industrie für Informationsverarbeitung* (in gedruckter und/oder elektronischer Form) wandeln müssen, wenn sie sich nicht von der Weiterentwicklung der Informationstechnologien abkapseln will." /Teichmann u.a. 1982, 23/

Für das Elektronische Publizieren sind in erster Linie zwei Entwicklungen von Bedeutung:

1. Die elektronische Texterfassung und die Satzvorbereitung finden zunehmend außerhalb der Druckbetriebe (Stichworte: Lohnsatz, Autorenerfassung) statt. Durch die Fremddatenübernahme kann die Neuerfassung von Manuskripten entfallen, und die Erfassungskosten reduzieren sich. Wenn die Manuskripte bereits ausgezeichnet sind, kann zudem ein Teil der Kosten für die (maschinengerechte) typographische Aufbereitung der Manuskripte eingespart werden.
2. Die einmal maschinenlesbar vorliegenden Daten bzw. das einmal erstellte Satzband können - elektronische Ausgabe- und Lieferformen vor Augen - mehrfach genutzt werden. Mit Mehrfachnutzung sind sowohl die Ausgabe auf Druckern verschiedener Leistung als auch die Nutzung des Satzbandes für den Datenbankaufbau, die Einspielung von Satz in Btx-Angebote oder auf Bildplatten u.ä. (multi-media-publishing) gemeint.

Die Herausforderung der neuen Technologien darf aber nicht als Alternative Papier oder Elektronik mißverstanden werden. Er wird oft übersehen, daß mit elektronischen Kommunikationsmitteln auch

ein Anwachsen von Papierprodukten verbunden sein kann. Altbekannte Beispiele sind die Fernsehzeitschrift, das Buch zum Film, die Computerzeitschrift oder die umfänglichen Dokumentationen zu Softwareprodukten.¹³

Andererseits ist nicht unwahrscheinlich, daß die Bedeutung des Papiers als Archivmedium zurückgehen wird /vgl. Strassmann 1985, 351/ und etwa Microfilm (bereits heute) und Bildplatte seine Stelle einnehmen.

Es ist auch nicht unwahrscheinlich, daß der wichtige Bereich der Werbung sich verstärkt den elektronischen Medien zuwendet, aber auch das bedeutet nicht unbedingt die Verdrängung des Papiers.

"Es wird den Printmedien bestimmt nicht gelingen, den derzeitigen Anteil von 75 Prozent an den Werbeausgaben zu halten. Da aber die Werbeausgaben zunehmen werden und außerdem die neuen Medien auch 'Kunden' der Printmedien sein werden, geht die Papierindustrie davon aus, daß der Bedarf für graphische Papiere und damit auch der Markt für die Printmedien weiter zunehmen wird." /Schrank 1986, 932/

Andererseits ist wiederum richtig, daß das Papier die "Achillesferse der gedruckten Medien" /Radzke 1984, 327/ ist. Damit meint Radzke, daß sowohl die Rohstofffrage¹⁴ als auch die Energiekostenfrage bei der Papierproduktion sich längerfristig negativ auf die Printmedien und ihren Markt auswirken werden, wodurch im Gegenzug elektronische Angebote profitieren könnten - nicht unbedingt, weil sie besser, sondern weil sie billiger sind.

Was das Elektronische Publizieren angeht, macht schon ein grober Blick auf den Ist-Zustand deutlich, daß das elektronische Angebot

¹³ Ein weiteres Beispiel: auf der INFOBASE 1986 wurde ein Vortrag von computererzeugten Folien begleitet, und weil der Vortrag ganz gut ankam, verschickte der Referent an alle Zuhörer (ca. 200) seinen "Folienvortrag" (ca. 30 Seiten) auf Papier. Das Schlagwort von der "papierlosen" Gesellschaft wird gerne auf Lancaster /Lancaster, 1978/ zurückgeführt, der es zwar benutzt /vgl. S.153/, aber nicht zentral. Ihn interessieren hauptsächlich Informationssysteme im technisch-wissenschaftlichen Bereich.

¹⁴ Das Durchforstholz, Grundstoff der Papierindustrie, wird z.B. wegen des zunehmenden Weitpflanzverbands in Zukunft immer weniger im Holzangebot auftauchen. /Vgl. Radzke 1984, S.328/

an Publikationen, die nicht gleichzeitig auch als Druckerzeugnisse vorliegen, vorerst noch gering ist.

2.2.3.2 Verlegerstandpunkt

Eine andere Sicht des Elektronischen Publizierens ergibt sich, wenn der Verlegerstandpunkt eingenommen wird. Für den Verleger ist allerdings wie für den Setzbetrieb durch die Telematik eine neue Situation geschaffen.

Der Verlag hat zunehmend mit Autoren zu tun, die ihre Manuskripte maschinenlesbar einreichen. Vorteile für Autor und Verlag können primär darin gesehen werden, daß ein Elektronisches Manuskript der Öffentlichkeit schneller zugänglich gemacht werden kann, die Informationen also aktueller und kostengünstiger sind, da ja die Neuerfassung der Texte (im günstigen Fall) nicht mehr nötig ist. Fehler, die durch eine Neuerfassung neu entstehen, könnten werden vermieden. Auch diese Manuskripte müssen lektoriert und redigiert werden, bevor sie gesetzt werden können. Die Elektronischen Manuskripte führen zu neuartigen Bearbeitungsverfahren und neuartigen Schnittstellen zwischen Autor und Verlag einerseits und Verlag - Druck andererseits.

Da ein Verleger ja "geistiges Eigentum" vermarktet, ist er, ebenso wie der Druckbereich, an Mehrfachnutzung interessiert. Mehrfachnutzung beginnt damit, einmal erfaßte Daten für Neuauflagen oder veränderte Auflagen wiederzuverwenden; sie bedeutet, aus einem Datenbestand mehrere Produkte zu gewinnen (Artikel, die in verschiedenen Zeitschriften erschienen sind, können zu Sammelbänden kompiliert werden; aus den Abstracts und bibliographischen Angaben von Aufsätzen kann ein Referatedienst erstellt werden; es können Lang- und Kurzfassungen erstellt werden etc.) und bedeutet last not least, Daten über verschiedene Kanäle ("Medien"), also auch elektronisch, anzubieten. An dieser Stelle wird der Verleger zum elektronischen Verleger, zum "data-base-publisher", "multi-media-publisher" u.ä.

Die Perspektive des Verlegers dreht sich, und das macht ihre Spezifik aus, um das Veröffentlichen auf Bestellung, "publishing on demand", und die elektronische Dokumentversorgung, "document delivery". Im ersten Fall, dem Veröffentlichen auf Bestellung (von der anderen Seite Wahlkommunikation), spart der Verleger Lagerkosten ein und reduziert sein Risiko, auf hohen Auflagen "sitzen zu bleiben". Das, was er liefert, kann ein Druckerzeugnis oder eine elektronische Publikation sein. Das "publishing on demand" kann auch als Ergänzung zum Online-Ordering gesehen werden. Die bestellten Artikel würden dann nicht mehr als Kopien von einer Bibliothek (was

dem Verlag nichts einbringt), sondern als Originalartikel vom Verlag geliefert.

Die elektronische Dokumentlieferung hat außer mit dem Erschließen neuer Märkte auch mit dem Rückgewinnen verlorenen Terrains zu tun. Das "Kopierunwesen" wird dafür verantwortlich gemacht, daß kleinere Fachzeitschriften nicht mehr die nötigen Auflagen erreichen, um sich zu tragen, und folglich eingestellt werden mußten.¹⁵ Ein weiterer Grund für den Problemdruck bei den Verlagen geht auf die gekürzten Bibliotheksetats zurück.

Werden nun diese Texte im Rahmen einer Volltextdatenbank¹⁶ angeboten, ergibt sich ein anderes Kalkül, weil einerseits eine große Datenmenge vorgehalten wird, andererseits aber nur für jeden abgerufenen Artikel gezahlt wird. Geht man davon aus, daß die Druckversionen nicht parallel zur Verfügung stehen, folgt, daß das Kopieren von Zeitschriftenartikeln drastisch zurückgehen wird.

Die elektronische Dokumentauslieferung kann also als ein Schritt gesehen werden, schnell "entlegenste" Artikel an entlegenste Orte zu liefern; sie bedeutet aber auch implizit, daß Verleger die marktmäßige Organisation eines Ausschnitts der "Wahlkommunikation" betreiben oder, weicher formuliert, ihre Aufgabenteilung mit den Bibliotheken neu bestimmen müssen.¹⁷ Hugh E. Look kritisiert den Verlegerstandpunkt, weil das Konzept des "Document delivery" noch an dem Vorstellungshorizont der Printmedien gebunden bleibe.

¹⁵ "Viele wissenschaftliche Fachzeitschriften sind in den letzten Jahren in ihren Auflagen zurückgegangen. Ein Grund hierfür liegt darin, daß von einem Exemplar immer häufiger Kopien gezogen werden" /Teichmann u.a. 1982, 232/.

¹⁶ Die Volltextdatenbank bietet in der Regel nicht den ganzen Artikel, sondern nur die zeichencodierten Textelemente, was den Wert der Volltextdatenbanken natürlich einschränkt und andererseits Entwicklungen zum Angebot auch der Graphiken, Formeln u.a. komplexer Textelemente anregt hat. Manfred Krüger spricht in bezug auf die sogenannten Volltextdatenbanken vom "unvollständigen Volltext" /vgl. Krüger 1986a, 150/ .

¹⁷ "... a reasonable and understanding dialogue must be maintained between publishers and librarians. We therefore urge that ... the two professions regard themselves as complementary in these new enterprises" zitiert aus einer gemeinsamen Stellungnahme europäischer Bibliothekare und Verleger /Impact Electronic Technology 1985, 8/.

"The concept of text and graphics as displayed on a page - the printed image - is paramount and all the skills are employed to keep this essential feature. Document delivery systems are, therefore, designed to match our 'inbred' needs for books and print generally. They benefit from a pre-existing market and preserve the status quo." /Look 1983, 163f/

Es fragt sich allerdings, ob mit dieser Orientierung am "printed image" nicht auch Qualitätsstandards verbunden sind, Leistungsmerkmale der Printmedien, die mit Recht hochgehalten werden sollten. Die Drucktechnologie hat z.B. Standards an typographischer Qualität, Integration von Text und Bild gesetzt und die Publikationsformen Buch und Zeitschrift z.B. Standards der Handhabbarkeit, die einen Maßstab abgeben, an dem sich elektronische Lieferformen und Endgeräte messen lassen müssen.

Köhler u.a. /Köhler u.a. 1986/ zeigen am Beispiel Textverarbeitung und Dokumenterstellung, wie typographisches Wissen in Computerprogrammen und -systemen geradezu mißachtet wird. Einige Beispiele:

- "Das Verhältnis zwischen Schriftgröße und Durchschuß - für das leichte Lesen besonders subtil - wird völlig mißachtet" /Köhler u.a. 1986, 4/.
- Typographisches Wissen um Schriftbilder, z.B. "Unterschneiden" wird nicht in die Programme eingebracht /vgl. ebd./.
- Der Umbruch ist in der "schwarzen Kunst" an Absätzen orientiert, in "normalen" Textverarbeitungssystemen an Zeilen.
- Die Aus- und Eingabe einer DIN-A4- Seite sollte auch auf einem Sichtgerät, das DIN A4 darstellen kann, erfolgen
- Die Gestaltung einer Bildschirmseite bei der Eingabe wie bei der Ausgabe sollte in etwa der Gestaltung von Druckseiten entsprechen.

Was das Elektronische Publizieren angeht, macht schon ein grober Blick auf den Ist-Zustand deutlich, daß das elektronische Angebot an Publikationen, die nicht gleichzeitig auch als Druckerzeugnisse vorliegen, vorerst noch gering ist.

Das Verhältnis elektronischer zu gedruckten Publikationen ist also nicht als Substitutionskonkurrenz zu verstehen; oftmals wirbt das elektronische Produkt sogar für das Printprodukt.

Als Illustration, da hier nicht das Gesamtangebot elektronischer Angebote systematisiert und dargestellt werden kann, folgt umseitig

eine kurze Zusammenstellung von elektronischen Angeboten zum Elektronischen Publizieren.

Der Blick auf diese Übersicht zeigt, daß die Elektronischen Publikationen zum Elektronischen Publizieren bereits eine Reihe von Volltextangeboten aufweisen, die parallel angeboten werden, in einem Fall sogar noch auf CD-ROM. Darüber hinaus ist abzulesen, daß bereits Publikationen angeboten werden, die *nur* elektronisch zu haben sind.

Ein weiteres läßt sich noch aus dieser Tabelle ersehen, daß es schwerpunktmäßig Nachschlagwerke und Newsletter sind, die sich als Formen eignen.¹⁸ Daneben finden sich aber auch Angebote im Experimentierstadium, die innovative Züge tragen, z.B. eine direkte Eingabe von Autorenartikeln gestatten, ohne den Weg über ein Printprodukt.¹⁹

Es läßt sich vermuten, daß in einer ersten Phase das Parallel-Publizieren, d.h. zu einem Druckerzeugnis existiert ein elektronisches Pendant, vorherrschen wird und erst nach und nach ausschließlich elektronische Angebote auftauchen, die Druckversionen zunächst nachahmen, in einer weiteren Phase sich aber zunehmend davon unterscheiden werden durch spezifische Leistungsmerkmale.

2.2.3.3 Standpunkt der Informationsindustrie

Der dritte Standort, von dem aus Elektronisches Publizieren thematisiert wird, ist der der (elektronischen) Informationsindustrie. Ihr Herangehen kappt in gewissem Sinne den Traditionsstrang zum Druckerzeugnis. Ihr Denken ist nicht produkt- sondern dienstleistungsorientiert. Information wird als Ware, die sich rechnen und ihren Markt finden muß, weniger als Kulturgut, gedacht.²⁰ Look vertritt diese Position:

¹⁸ Vgl. zu den elektronischen Nachschlagewerken /Cook 1984/, /Johnson, Kutz 1984/ und /Roth, S. 1985b/; zu dem Host Newsnet, der nur newsletter anbietet /Simpson 1984/.

¹⁹ /Collier 1984/ beschreibt das "Electronic Magazine". Andere Experimente mit "elektronischen Zeitschriften" beschreiben /Gabel-Becker, Loeben 1986/, darunter auch das englische BLEND-Projekt, das über seine publikationsfreudigen Autoren Pullinger und Shackel recht bekannt geworden ist. Übersichten zu den Experimenten der EG in diesem Zusammenhang finden sich vornehmlich in den Artikeln von Mastroddi und Vernimb (siehe Literaturverzeichnis).

²⁰ Anzumerken ist hier, daß es vor allem Wirtschaftsinformationen

Informationsbanken zum elektronischen Publizieren

Name	Produzent	Anbieter	Bemerkungen
INSPEC	Institution of Electronic Engineers	DIALOG INKA DATASTAR	bibliographisch; parallel
LISA	Library Association Publishing	DIALOG	bibliographisch; parallel
ISA	IFI/Plenum Data Company	DIALOG	bibliographisch; parallel
EPA	PIRA	PERGAMON- INFOLINE	bibliographisch; parallel
(FOGRA)	FOGRA	ODAV	bibilographisch
INFODATA	GID	GID	bibliographisch; parallel
COMPUTER DATA BASE	Information Access Company (IAC)	DIALOG DATASTAR	bibliographisch; <u>Volltext</u> - Inhaltsverz.
ONLINE CHRONICAL	Online Inc.	DIALOG	<u>Volltext</u> ; z.T. nur online
DATABASE OF DATABASES	Martha Williams/Elsevier	DIALOG	Datenbankführer
.MENU-ISD	International Software Database Corporation	DIALOG	Online-ordering von Software möglich
MAGAZINE ASAP	Information Access Company	DIALOG	65 Volltextmagazine <u>Information Publishing</u> online, Rest parallel
ACADEMIC AM. ENCYCLOPEDIA	Grolier Electronic Publishing Inc.	DIALOG DATASTAR	auch als CD-ROM
DIRECTORY OF ONLINE DATABASES	Cuadra	DATASTAR	Datenbankführer parallel
ELECTRONIC MAGAZINE	Learned Information	ESA/IRS	<u>Volltext</u> ; Test; nur online
COMPUTER COMPACTS	North-Holland	INKA (?)	<u>Volltext</u> ; parallel
EURONET DIANE Verz.	EG	ECHO	Datenbankführer; kostenlos; parallel
ELECTRONIC INFORMATION REPORT	Link Resources	NEWSNET	Newsletter; <u>Volltext</u>
DATA BASE INFORMER		NEWSNET	Newsletter; <u>Volltext</u>

Abbildung 2. Informationsbanken zum Elektronischen Publizieren.

"We can distinguish electronic publishing as a new medium for presenting information to the user. The medium allows new approaches to be made to the presentation of material with combinations of text, sound, etc. to convey the information. The author and publisher of the material do not necessarily think in terms of a printed page image, and there may be no recognisable printed product." /Look 1983, 164/

Von dieser Warte aus steht nicht die um das "Medium" Papier etablierte, teils öffentlich-institutionelle, teils marktorientierte, Publikationskette im Vordergrund, sondern ein Zusammengehen verschiedener Industriezweige: Hardwarehersteller, Softwarefirmen, Netzbetreiber und Informationsanbieter bei der Erbringung von Informationsdienstleistungen.

"The consequence of the need for these four (Hardware, Software, Netze, Information - K.B) very different components in electronic publishing is that any one service may require resources from a good number of other firms, and in many cases those other firms are part of formerly quite independent industries"

/Hart, Rice 1985, 201/.²¹

Die sozio-technische Figuration des Elektronischen Publizierens ist hier in den informationsindustriellen Komplex eingelassen. Der Verleger ist nur noch einer unter mehreren Informationsanbietern.²² Aus der Sicht der Informationsindustrie

sind, die sich rechnen, noch weit vor wissenschaftlichen Daten aller Art. Aktuelle Informationen bieten /Silverstein, Ellwell 1985/ und /Irwing 1986/. Gisela Roth /Roth, G. 1986, 109/ bemerkt dazu in einem Bericht von der EUSIDIC-Jahrestagung 1985: "Es ist interessant festzustellen, daß die Online-Informationsindustrie (Textinformation) weltweit nicht einmal 50% der Umsätze eines einzigen Unternehmens wie Reuters erreicht. Es ist fraglich, ob diese Industrie jemals in die 'Bundesliga' der Informationsindustrie (im weiteren Sinne) aufsteigen kann". Skepsis gegenüber den Onlinediensten wird auch in /Schulte-Hillen, Wietersheim 1984/ verbreitet.

²¹ Diese Sicht teilt auch Gibbins /1984 a / und /1984 b/.

²² Das bemerkt Look nicht ohne Biß: "When the traditional publishers arrive on the electronic publishing scene, they soon discover that the club already has other members and that the language and vocabulary are strange. Publishers have no hope of taking over electronic publishing, but they do have a wide

kann vielleicht noch am ehesten eingebracht werden, was "telematische" Dienstleistungen von Printmedien positiv unterscheiden kann:

- daß die Dienstleistungen interaktiv ausgelegt sein können (z.B., daß Voraussetzungen direkter elektronischer Kommunikation zwischen Akteuren geschaffen werden, die diesen ermöglicht "ihre" Informationen auszutauschen.)
- daß die Dienstleistungen auf Weiterverarbeitbarkeit ausgelegt sein können (z.B., daß eine nachrichtentechnisch übertragene Graphik einer Schraube direkt in einem CAD-System weiterverwandelt werden kann)

Diese Beispiele machen nochmals deutlich, daß "Elektronisches Publizieren" nur zum Teil darin aufgeht, Druckerschwärze in elektronische Schläuche umzufüllen.

Die Präsentation der Standorte diene dazu, interessen geleitete Vorverständnisse herauszupräparieren, die der Publikationskette und dem Hypothesenraster etwas "Fleisch" verleihen sollten. Im folgenden geht es darum, in Hinblick auf den Gang dieser Arbeit, einen operationalen Zugang zum Elektronischen Publizieren zu gewinnen, der auch die Erörterung technischer Fragen gestattet.

2.3 ELEKTRONISCHES PUBLIZIEREN IM RAHMEN DER TEXTKOMMUNIKATION

Es erscheint praktikabel, Elektronisches Publizieren durch einen weiten Begriff von Textkommunikation zu bestimmen:

"Textkommunikation(in einem engen Sinne - K.B.). Sammelbegriff für die nachrichtentechnische Übermittlung von Text in zeichencodierter Form. Die Codierung besteht darin, daß jedem Textzeichen umkehrbar eindeutig ein Signal (meist ein Digitalsignal) zugeordnet ist. Man unterscheidet bei der Textkommunikation

- überwiegend papiergebundene Formen...
- überwiegend bildschirmgebundene Formen... . . .

and usefull variety of skills and expertise to bring to the field." /Look 1983, 175/

Häufig rechnet man auch die Faksimileübertragung ... zur Textkommunikation, obwohl sie eine Form der Festbildkommunikation ist." /Fellbaum Hartlep 1985, 248/

In dieser Definition, die geschrieben wurde, um Telex, Teletex, Videotex und Videotext unter einen Begriff zu bringen, fehlt Elektronisches Publizieren offensichtlich. Für den Bezug auf Elektronisches Publizieren sind noch Differenzierungen in drei Richtungen hinzuzufügen.

1. Textkommunikation sollte vorderhand nicht auf nachrichtentechnische Übermittlung beschränkt werden.
2. Der Bezug zur "systematisch offenen Kommunikation" muß noch eingebracht werden, *und*
3. die Darstellungsform Text muß noch weiter differenziert werden.

ad 1

Texte (schriftliche Fixierungen zunächst) können transportiert werden. Der konventionelle Bereich, etwa des Briefeschreibens und des Transports mittels der Post, sollte nicht aus dem Bereich der Textkommunikation herausdefiniert werden. Texte können aber auch codiert werden, um entweder nachrichtentechnisch übermittelt oder auf Datenträgern transportiert zu werden - auch in dem letzten Fall kann von Textkommunikation die Rede sein. Auch transportorientierte Kommunikationsmittel wie Bildplatte und Diskette können für das Elektronische Publizieren eingesetzt werden. Das wäre die erste Erweiterung der obigen Definition.

ad 2

Kommunikationsmittel können danach unterschieden werden, in welchen Kommunikationsbereichen sie eingesetzt werden können. Wersig /1985 a, 124f/ unterscheidet, jenseits der Individualkommunikation, Gruppen-, Wahl- und Massenkommunikation. Die Wahlkommunikation ist dabei ein Komplement zur Massenkommunikation, setzt Sammlung von Angeboten und den *wahlfreien Zugang* darauf voraus. Wahl- und Massenkommunikation sind öffentlich. Die folgende Abbildung macht deutlich, wo das Elektronische Publizieren im Rahmen der Textkommunikation verortet ist.

Diese Abbildung ist ein Ausschnitt einer größeren Abbildung, der aus Wersig /Wersig 1985a, 165/ übernommen wurde. Hier wird deutlich, daß Elektronisches Publizieren im Bereich der Massenkommunikation und Wahlkommunikation anzusiedeln ist.

Übertragung	Gruppenkomm.	Wahlkomm.	Massenkomm.
energetisch	Teletext	Bildschirmtext	Kabeltext
	Bildschirmtext	Kabeltext	Videotext
	Telekonferenz	electronic	electronic
	LAN	publishing	publishing

transport-orientiert	Vervielfältigung	Bibliothek	Print-Medien
	Bildplatte		Bildplatte

Abbildung 3. Ort des Elektronischen Publizierens I

Für die Zwecke dieser Arbeit können einige Modifizierungen an dieser Übersicht angebracht werden.

Nicht deutlich wird, daß Elektronisches Publizieren auch in den Bereich der Gruppenkommunikation hineinspielt. Die "elektronische Zeitschrift" wäre z.B. die Nutzung von Telekonferenzsystemen, um so etwas wie eine Fachzeitschrift elektronisch nachzubilden. Andererseits "verstößt" ein Telekonferenzsystem gegen das Kriterium Öffentlichkeit, die sich durch den allgemeinen Zugang definiert.

Es ist auch zu überlegen, ob Bildschirmtext, Kabeltext und Videotext, d.h. der ganze Bereich der Textkommunikation, der auf größere Publika aus ist, zum Elektronischen Publizieren gezählt werden sollte. Für Neustadt in "The Birth of Electronic Publishing" ist das keine Frage:

"The concept of electronic publishing is simple: pages of text and graphics are displayed on a television set or other inexpensive screen." /Neustadt, 1982, 5/

Im Prinzip scheint mir Elektronisches Publizieren als Oberbegriff zulässig, allerdings fehlt dann an anderer Stelle ein spezifischer Begriff für die Aktivitäten, die sich darauf richten, das Leistungs- und Angebotsspektrum der Printmedien in elektronischen Formen ab- und nachzubilden. Die Hauptaktivitäten in diesem Feld fallen in den Bereich der Fachkommunikation und beziehen sich dort auf das Angebot von Volltexten, die z.T. noch Graphik, Formeln, Tabellen und Abbildungen integrieren.

Nach diesem Verständnis ist Elektronischem Publizieren ein Platz neben anderen Angeboten, wie in der Abbildung (oben), einzuräumen.

Der Bereich der Bürokommunikation als Sonderfall der Gruppenkommunikation kann aus der Betrachtung ausgeklammert werden - auch wenn die technischen Entwicklungen in diesem Gebiet nicht ohne Bedeutung für den Bereich öffentlicher Kommunikation sind.²³

Insgesamt ergibt sich also ein modifiziertes Bild (vgl. Abbildung 4 auf Seite 27), in dem Elektronischem Publizieren eine übergeordnete Rolle zukommt. Elektronisches Publizieren ist demnach nicht an bestimmte Technologien gebunden und nicht an bestimmte existierende Dienste. Entscheidend ist, ob die Technologien und Dienste zur Textkommunikation im öffentlichen Raum benutzt werden. Damit ist Elektronisches Publizieren prinzipiell offen für neue technische Entwicklungen konzipiert und fällt nicht mit bestimmten Kommunikationsmitteln zusammen. Es unterscheidet sich technisch von konventioneller Textkommunikation entweder durch die Übertragungsweise (energetisch) und/oder durch die maschinenlesbare Speicherform der zu kommunizierenden Textinformationen.²⁴

ad 3

Ich will die verschiedenen Textelemente, die bei Elektronischem Publizieren vorrangig in Betracht kommen, schon an dieser Stelle kurz beschreiben, da sie für die Komplexität der anstehenden Probleme und technischen Lösungen maßgeblich sind.

Ein Text kann aus mehr oder weniger komplexen "Textelementen" bestehen:

1. "Text ist dadurch gekennzeichnet, daß die Zeichen aus einem vereinbarten Alphabet linear angeordnet sind und zu ganz bestimmten Strukturen zusammengefaßt werden." /Petersen 1985, 162/ Zeilenanordnung, Satzspiegel, Abstände, Hervorhebungen, Schriftarten und -größen u.a. bestimmen die Gestaltung der Strukturen.
2. "Tabellen sind in der Regel spalten- und zeilenstrukturiert angeordnete Textelemente, deren Struktur durch Linienmuster oder durch Hinterlegen von Rastern hervorgehoben wird. Ihre

²³ Vgl. zur Telematik im Bürobereich /Rauch 1983/ und zu den sich auflösenden Grenzen zum Publikationsbereich /Blake 1982/.

²⁴ Wenn die Faksimilezeitung einmal Wirklichkeit wird und die Technik nicht nur zum Verschicken von Zeitungsseiten von Redaktionen zu weit entlegenen Druckereien eingesetzt wird, ist sie selbstverständlich eine Praxis des Elektronischen Publizierens. Ähnliches ließe sich über den Einsatz der CD-ROM sagen.

Übertragung	Gruppenkomm.	Fachkommunikation	
		Wahlkomm.	Massenkomm.
Elektronisches Publizieren			
energetisch	"elektronische Zeitschrift"	Bildschirmtext	Kabeltext
	Telekonferenz	Kabeltext	Videotext
	Bildschirmtext	Online-Dienst	
	LAN	Volltexte + Graphik	Volltexte + Graphik

	Bildplatte	Host	Bildplatte Diskette
transport-orientiert		konventionelles Publizieren	
	Vervielfältigung	Bibliothek	Print-Medien

Abbildung 4. Ort der Elektronischen Publizierens II

Funktion ist die möglichst übersichtliche Darstellung von Zusammenhängen und Ordnungen."/ebd./

3. Formeln. "Generell versteht man unter Formeln die zweidimensionale Anordnung von Zeichen. Je nach Anwendungsgebiet sind ganz bestimmte Strukturen üblich, etwa chemische Formeln oder mathematische Formeln. Die Darstellung mathematischer Formeln z.B. erfordert einen erweiterten Zeichensatz..."/ebd/
4. Rechnererzeugte und scannererfaßte Graphiken. Rechnererzeugte Graphiken werden mit Sprachmitteln beschrieben und können vor der Erzeugung der Graphik in einem Metafile vorgehalten werden. "Gescannte" Graphiken liegen digitalisiert als bitmap vor und sind mit anderen Mitteln zu bearbeiten.
5. Grautonbilder sind in der Regel auch scannererfaßt.

Grundsätzlich wichtig ist die Unterscheidung von zeichencodierten und bitorientierten Speicher- und Darstellungsformen. Innerhalb der zeichencodierten Textelemente stellen Standard- und erweiterte

Zeichensätze eine praktisch wichtige Einteilung dar.²⁵ Nur wenn man bereit ist, den Begriff Text unabhängig von einer bestimmten Darstellungs- und Speicherform anzusetzen, ist man in der Lage, das gesamte Textmaterial, das üblicherweise in Publikationen vorfindbar ist - auch Farbe z.B. - zu berücksichtigen. Ich komme damit zu einer vorläufigen Definition des Elektronischen Publizierens, die als "angepaßte" Version der eingangs des Kapitels zitierten gelten kann.

Unter Elektronischem Publizieren sollen systematisch offene Formen der Textkommunikation (i.w.S.) über anerkannte Kanäle verstanden werden, die zu ihrer Realisierung elektronischer Hilfsmittel (Hard- und Software und/oder Netze) bedürfen und für den zeitpunkt-unabhängigen Gebrauch geeignet sind.²⁶

Eine terminologische Schwierigkeit wird im weiteren Text, hoffe ich, keine Verständnisprobleme bereiten.

Auf die Frage, was ein Text, was ein Dokument, was ein Manuskript, was folglich Textelemente und was Dokumentelemente sind, läßt sich keine einfache Antwort geben; es würde desweiteren auch nichts an dem "unabgestimmten" Gebrauch ändern. Ich setze darauf, daß aus dem jeweiligen Kontext - Dokumenterstellungsprozeß, Dokumentarchitektur, Publikationsprodukt, nachrichtentechnische Übertragung - klar wird, wie die Konzepte jeweils verwendet werden.

²⁵ In der Definition von Fellbaum, Hartlep waren bereits beide Darstellungsformen angesprochen.

²⁶ Es wird natürlich niemand aufgefordert, nun seiner Definition abzuschwören, die vielleicht nicht-öffentliche Kommunikation dazu oder nicht-netzgebundene elektronische Kommunikationsmittel nicht dazurechnet etc. Es wäre auch möglich, verschiedene Definitionen für verschiedene Bereiche nebeneinander bestehen zu lassen.

3. VORÜBERLEGUNGEN ZUR SGML

In diesem Kapitel werden Vorüberlegungen zum Elektronischen Publizieren mit der Standard Generalized Markup Language vorgelegt. Dabei werden unterschiedliche Hinführungen aufgewiesen.

In der ersten wird unter dem Titel Autor - Verlag - Satz die neue Situation, die durch Elektronische Manuskripte entstanden ist, geklärt und auf den Punkt zugespitzt, daß der Einbezug des Autors in den technischen Produktionsprozeß recht eigentlich da beginnt, wo er anfängt, seinen Text "auszuzeichnen". Folgerungen für den Satzbereich werden miterörtert.

In der nächsten Hinführung wird ein allgemeines Modell des Dokumenterstellungsprozesses auf dem ISO-Schichtenmodell aufbauend entwickelt, in dem ein Elektronisches Manuskript in ein systematisches Verhältnis zu späteren Bearbeitungsstufen gesetzt wird und komplexe Textelemente berücksichtigt werden. Damit soll ein Bezugsrahmen zur Erörterung technischer Probleme im Umfeld der SGML geliefert werden.

In der dritten Hinführung wird die Entwicklung der SGML aus einer Kritik herkömmlicher Textverarbeitungssysteme für Zwecke des Elektronischen Publizierens hergeleitet.

3.1 DIE BEZIEHUNG AUTOR - VERLAG - SATZBETRIEB

Die Autor-Verlags-Beziehung unter dem Einfluß der neuen Informations- und Kommunikationstechniken ist ein kontrovers diskutiertes Thema. Fabian /1983, 289/ zählt die Funktionsverschiebungen zwischen Autor und Verlag zu den "gravierendsten Veränderungen", die die neuen Technologien in dem Bereich bewirken. Man sollte hinzufügen, daß auch für den Satzbereich gravierende Veränderungen an der Tagesordnung sind. In diesem Kapitel wird die Publikationskette in dem Input-Bereich bis zur "Datenbasis", wie sie in Abbildung 1 auf Seite 8 dargestellt ist, genauer untersucht.

3.1.1 Das Rationalisierungspotential

Ein Blick auf die "traditionelle" Beziehung von Autor und Verlag macht klar, an welchen Punkten Veränderungen ansetzen konnten und können:

1. Ein Autor erstellt ein Manuskript handschriftlich oder mit der Schreibmaschine und schickt es an einen Verlag.
2. Das Lektorat des Verlags prüft die Qualität des Manuskripts und nimmt es - in dem Fall, der uns interessiert - an, nicht ohne dem Verfasser inhaltliche und stilistische Änderungsvorschläge zu unterbreiten, die dann im Verlag oder vom Autor ganz oder teilweise in das Manuskript eingearbeitet werden. Diese Autor - Verlagskommunikation kann mehrere Durchläufe umfassen.
3. Wenn das Manuskript "druckreif" ist, wird es für den Satz aufbereitet. In der Satzvorbereitung wird ein Manuskript zunächst ausgezeichnet, mit Mark-up²⁷ versehen, d.h. mit Markierungen, Satzanweisungen.
4. Diese Satzanweisungen hatte der Setzer beim Handsatz direkt umzusetzen. Beim Maschinensatz werden zusammen mit der Erfassung des Textes die Auszeichnungen auf dem Papier in Steuerbefehle für die Maschine umgesetzt, indem entsprechende Codes in den Text eingestreut werden.
5. Wenn wir den Fotosatz weiterverfolgen,²⁸ wird jetzt der Text auf der Fotosatzanlage belichtet. Ergebnis sind die Korrekturfahnen.
6. Diese werden in der Regel an Autor und Verlag geschickt, die Fehler, die bei der Erfassung des Manuskripts entstanden sind, suchen und nach den Korrekturvorschriften DIN 16511 am Seitenrand der Fahnen verbessern.
7. In der Setzerei werden die Korrekturen ausgeführt, der Text neu belichtet und die Seiten montiert.

²⁷ "Publishers are familiar with the concept of mark-up in the context of preparing of typed or other paper-based manuscripts for typesetting. The traditional approach involves the annotation of textual matter with **marks** which, by convention, indicate to the compositor what are the style and layout requirements of the document." /Cave 1984, 2/

²⁸ Für eine ins Detail gehende und verschiedene Arbeitsabläufe beim Fotosatz berücksichtigende Darstellung, die durch Ablaufdiagramme noch illustriert ist, siehe /Teichmann u.a. 1982, 97 - 108/.

8. Der Autor erhält die umbrochenen Seiten und gibt sein *Imprimatur*.

Eine entscheidende Rationalisierung ergibt sich, wenn die Neuerfassung entfällt. Wenn die Texte maschinenlesbar bereits beim Autor anfallen, werden Kosten und Zeitaufwand für die Neuerfassung gespart. In Folge fallen auch die Korrekturläufe, die sich auf Eingabefehler bei der Neuerfassung richten, weg. Auch hier werden nochmals Zeit und Kosten eingespart. Ein zusätzlicher - und im Zentrum dieser Arbeit stehender - Rationalisierungseffekt entsteht, wenn das Markup, das normalerweise an den Seitenrändern der Manuskripte angebracht wird, in das Elektronische Manuskript verlegt wird und diese Informationen automatisch in Formatierbefehle des Satzrechners umgesetzt werden können.

Für den *Autor* ist die Beschleunigung des Publikationsprozesses anzustreben, weil sein Beitrag dann aktueller und schneller am Markt ist. Davon profitiert auch der Verlag.

Für den *Verlag* ergeben sich aus der "elektronischen Datenbasis", die durch die "Elektronischen Manuskripte" aufgebaut wird, eine Vielzahl möglicher Vorteile, die in "Die ISO/DIS 8879 (SGML)" auf Seite 66 aufgegriffen werden. Auf der Ebene konventioneller Druckausgabe ist der Vorteil, der bei der Herausgabe weiterer, u.U. veränderter Editionen entsteht, besonders hervorhebenswert. Auch hier ist der springende Punkt, daß die Neuerfassung entfallen kann. Für die *Satzbetriebe* folgt aus der neuen Situation Umstellung auf Fremddatenübernahme und Rückgang der Erfassungstätigkeit; "da noch 85 Prozent des Satzvolumens in Form von traditionellen Schreibmaschinenmanuskripten angeliefert wird, besteht hier noch ein erheblicher Rationalisierungsspielraum" /Perspektiven 2 o.J., 61/.²⁹

Dieses Rationalisierungspotential ist "infrastrukturell" an die Verbreitung von Textverarbeitungssystemen bei Autoren gebunden. 60% der zu übernehmenden Fremddaten werden auf PCs erstellt /vgl. Perspektiven 2 o.J., 61/.³⁰ Ohne weiteres allerdings stellt sich die

²⁹ Man kann es auch so beschreiben: "Die Satzbetriebe müssen sich zunehmend auf ihre Qualitäten besinnen, die nicht in der Erfassung, sondern in der Gestaltung liegen. Erfassung sollte nur einen zusätzlichen Service darstellen" /Teichmann u.a. 1982, 225/.

³⁰ Ein Autor wird hier abstrakt als jemand gesehen, der als Urheber geistigen Eigentums dies zur Veröffentlichung bringt. Damit ist nicht gesagt, in welchem institutionellen Rahmen und mit wel-

Rationalisierung nicht ein, selbst wenn ein Autor über ein Textverarbeitungssystem verfügt.

"It is all very well and good for the manuscript to be in electronic form, but it's not very usefull if the author's machine is the only one that can read it" /Roth S., 1985 a, 116/.

Die Skala, die sich hier auftut, reicht von einem für den Verlag nicht lesbarem Elektronischen Manuskript bis zu einem vollständig für den Satz aufbereiteten Elektronischen Manuskript (einschließlich Markup), das direkt von der Fotosetzmaschine verarbeitet werden kann.

Dazwischen liegen verschiedenste Varianten, Kompatibilität zwischen "Autorensystemen" und "Verlagssystemen" herzustellen, und verschiedenste Vorstellungen, den Arbeitsablauf - einschließlich der Konvertierprozeduren - bis zum fertigen Satzprodukt neu aufzuteilen.

Je nachdem, welcher Punkt der Skala ins Auge gefaßt wird und welcher Standpunkt eingenommen wird, fallen die Stellungnahmen zum "Elektronischen Manuskript" höchst verschieden aus.

So berichtet z.B. Peter Ferris (John Wiley & Sons Limited) von einem Satzexperten, der sagte "that if one of his students came to him with a manuscript on a floppy disk for publication, he told him to throw the disc in the bin and send the manuscript to Singapore to be typeset" /Ferris 1983, 57/.

Borko, der einige Verleger angeschrieben hat, ob sie sein mit WordStar auf einem IBM PC erstelltes Elektronisches Manuskript übernehmen wollten, zitiert aus einem Antwortbrief: "'It is this necessary encoding or recoding that makes it cheaper today to re-keyboard straight text than to process an input disk'" und fährt fort: "The significant issue raised in this publisher's letter is whether typesetting from author-produced diskettes can be made cost-effective" /Borko 1985, 3/.

Diese Zitate sprechen für die verbreitete These, daß der *Pull* von den Autoren kommt.³¹ Michael Schlecht von der IG Druck und Papier stellt die Gegenposition heraus:

cher technischen Ausstattung und personellen Unterstützung die Manuskripte erstellt werden.

³¹ Okeshott und White zitieren einige Erhebungen dazu. Eine Untersuchung der University Microfilms Inc. kam zu dem Ergebnis,

"In immer stärkerem Maße wird von den Unternehmern nach Wegen zur Verwirklichung der 'Einmaltexterfassung' gesucht. Besonderen Auftrieb hat dies erhalten mit der Verbreitung der *Personalcomputer (PC)*. ... Zum Beispiel gibt es Buchverlage, die von ihren Autoren Schreibmaschinenmanuskripte nicht mehr abnehmen" /Perspektiven 2, o.J., 60/.

Die anstehenden Fragen betreffen einmal die direkte Autor - Verlagsbeziehung, die sich mit dem Elektronischen Manuskript verändert, und zum anderen die Verschiebungen im Satzbereiches, die damit einhergehen. Davor jedoch ist darzustellen, wie die Übergabe respektive Übernahme Elektronischer Manuskripte sich technisch realisieren läßt.

3.1.2 Fremddatenübernahme

"Für den Betreiber eines Fotosetzsystems sind alle nicht direkt auf diesem System erfaßten Daten Fremddaten" /Knoth 1986a, 505/. Die Fremddatenübernahme umfaßt auch das Einlesen von Manuskripten über Blattleser oder anders auch OCR³² -Lesegeräte genannt. Diese Geräte, die anfangs nur die Normschriften OCR-A und OCR-B erkennen konnten, lesen heute ca. acht gängige Schreibmaschinenschriften. Diese Art der Datenübernahme wird auch durchaus von Verlagen eingesetzt /vgl. Ferris 1983, 58/. Auf dem Sektor der Schriftenerkennung gibt es außerdem die intelligente Lesemaschine von Kurzweil, einer Tochter von XEROX.³³ Das Problem der Fremddatenübernahme ist

daß mehr als die Hälfte der 30000 Doktorarbeiten, die jährlich in den USA entstehen, "elektronisch" erstellt werden. Oxford University Press schätzte, daß etwa 30% seiner Autoren "Floppy Disks" liefern könnten. Zu diesen und anderen Angaben siehe /Okeshott, White 1984, 15/. Vgl. auch die Ergebnisse einer Umfrage des Aspen Institute, zit. in /Sperling Martin 1985, 13-8f/, die in dieselbe Richtung gehen.

³² OCR=Optical Character Recognition

³³ Die Kurzweil Lesemaschine unterscheidet sich von den herkömmlichen Blattletern dadurch, daß sie neben fest installierten Zeichensätzen noch weitere dazulernen kann. Eine Seite eines Textes wird zeilenweise eingescannt. An einem angeschlossenen Bildschirm bildet das System das gescannte Pixelmuster des ersten Buchstaben ab und teilt mit, welches Zeichen es erkannt hat. Ist das Zeichen unrichtig identifiziert, kann ein Operator

in der Hauptsache aber ein Problem der Schnittstelle zwischen Textverarbeitungssystem und Satzrechner. Es besteht recht gesehen aus *drei* Inkompatibilitäten zwischen Textsystem und Setzsystem³⁴

1. unterschiedliche Datenträger
2. unterschiedliche Codes
3. unterschiedliche Formate

Relevante Datenträger in unserem Zusammenhang sind in erster Linie diverse Diskettentypen, die zu bestimmter Hardware gehören, - bestimmten Diskettenlaufwerken bestimmter Rechner. Bei einem Elektronischen Manuskript auf Diskette wäre zunächst zu fragen, ob der Satzrechner überhaupt Laufwerke für entweder 8", 5 1/4" oder 3 1/2" große Disketten hat.

Die Datenträger unterscheiden sich außerdem durch unterschiedliche Formatierung. Unterschiedliche *Betriebssysteme* speichern Informationen auf unterschiedliche Art und Weise auf Datenträgern. Ein Betriebssystem organisiert die Speicherung so, daß es auch wieder auf die Daten zugreifen kann. Datenträger werden zu dem Zweck "formatiert", d.h. es werden auf Disketten Sektoren (hard- oder softsektoriert) und Spuren festgelegt, Dichte und Anzahl der Spuren und ob eine Diskette ein- oder beidseitig zu "beschreiben" ist. Das Betriebssystem teilt dem Schreib-/Lesekopf des Laufwerks mit, wohin

das korrekte Zeichen zuweisen. Wenn alle Zeichen richtig identifiziert werden, kann das automatische Einlesen beginnen. Die Lesemaschine kann verschiedene Schriften und verschiedene Schriftgrößen auf einer Seite gleichzeitig verarbeiten /vgl. Information Systems Consultants 1985, 181-183/

³⁴ "Ein Textsystem, daß zu einem Setzsystem kompatibel ist, gibt es jedenfalls genauso wenig wie ein Setzsystem, das zu einem anderen Setzsystem kompatibel ist " /Knoth 1986a, 506/. In einer vielbeachteten Studie der Pira (Printing & Information Technology Division) wurde festgestellt: "Composition systems from different manufacturers work to completely different standards. There is no standardisation of either code structures, character sets, or typographic command structures. Even equipment from a single manufacturer is not necessarily consistent. A very important point is that the relationship between a code and a character is only defined within an individual installation of a composition system" /Pira 1982, 82f/. Diese Situation macht die Kompatibilität von Autoren- und Setzsystemen zu einer Sache der Absprache.

bestimmte Informationen geschrieben werden sollen.³⁵ Verschiedene Betriebssysteme bedeuten immer verschiedene Diskettenformate (MS-DOS vs. CP/M etc.). Aber auch ein und dasselbe Betriebssystem auf verschiedenen Rechnern führt nicht automatisch zu gleichen Diskettenformaten. Ein "ein bißchen" anderes Format heißt bereits, daß die Daten für andere Rechner schon nicht mehr vollständig lesbar sind.

Neben den unterschiedlichen Diskettenformaten werden z.T. auch ganz verschiedene Codes von verschiedenen Rechnern benutzt. Viele Mikrocomputer verwenden den ASCII-Code (American Standard Code for Information Interchange), IBM-Maschinen den EBCDIC (Extended Binary Coded Decimal Interchange Code). Der erste ist ein 7-Bit-Code, der zweite ein 8-Bit-Code. Im Satzbereich wird häufig ein 6-Bit-Code verwendet.

Eine weitere Schwierigkeit stellen formatierte Texte dar. Formatierung bezieht sich hier auf die Gestaltung des Textes. Mikrocomputer und Satzrechner sind in dieser Hinsicht verschieden komplex. Satzrechner kennen z.B. variable Abstände zwischen Buchstaben und variable Buchstabenweiten, Silbentrennprogramme und Ausschlußprogramme. Diese Eigenschaften, die die Unterschiede in der Leistungsfähigkeit zwischen Textverarbeitung und Satz beschreiben, werden zwar immer kleiner, das ist bei einer Datenübernahme aber kein Vorteil. Die Formatiercodes (für die Gestaltung), die eine Software für Textverarbeitung oder Satz verwendet, werden in der Regel mit keinem anderen System geteilt, nicht unter Mikrocomputersoftware und nicht unter Satzprogrammen.³⁶

Das kleinste Problem stellen die unterschiedlichen standardisierten Codes dar (ASCII, EBCDIC etc.). Ihre Unterschiede können über Konvertiertabellen, in denen korrespondierende Codes für ankommende und abgehende Zeichen ausgetauscht werden, behoben werden. Lediglich die Zeichen, die den Zeichensätzen nicht gemeinsam sind, müssen durch Ersatzzeichen dargestellt werden (die Abführungsstriche eines Satzrechners etwa durch "?").

³⁵ Das Laufwerk selbst kann u.U. die Möglichkeiten des Betriebssystems einschränken, wenn es z.B. nur einen Schreibkopf besitzt, kann es Disketten nicht zweiseitig beschreiben, oder wenn seine Kapazität auf 360K beschränkt ist, kann es keine 1,2 MB Disketten formatieren.

³⁶ Wer z.B. einen in WordStar erfaßten Text in Word aufruft, wird auf einen "merkwürdigen" Zeichensalat stoßen, der unterschiedlichen Formatiercodes geschuldet ist.

Für das Problem der Formatiercodes gibt es die Möglichkeit, die Steuerzeichen des Textverarbeitungssystems herauszufiltern und zu löschen.³⁷ Damit geht dann allerdings auch die Information, die sie enthalten, verloren.

"Wird uns nun solch ein vorab gestalteter Text zum Setzen und richtigen Drucken übergeben, wären wir doch reichlich töricht, wenn wir als erstes hingehen und aus ihm durch Löschen der Gestaltungsbefehle wieder endlosen Text machen" /Knoth 1986a, 506/.

Wenn man die Formatiercodes aber nicht verwenden kann oder will, ist es selbstverständlich anzuraten, daß Autoren ihren Text ohne Steuerzeichen abgeben, z.B. in reinem ASCII-Code.

Das Problem unterschiedlicher Speicherformate auf den Disketten kann bei Übernahme von Daten über Telekommunikationsnetze umgangen werden.

"... it is simpler to phototypeset text sent via the communications port of a word processor. The reason is that text information is stored in a simpler format for communications than for storage on floppy disk, or other magnetic media. This is due in part, to the greater degree of standardisation in telecommunications than in computing" /Holloway o.J., 12/.

Dieser Weg setzt entsprechende Kommunikationsfähigkeit der Rechner für die Telekommunikation voraus, ist aber auch eine Frage der Leitungskosten und der Datensicherheit.

Man stelle sich etwa die Übertragung eines Dreihundert-Seiten-Manuskripts über Telefonleitung mit 300 Baud von Berlin nach München vor. Es gibt aber auch die Möglichkeit, lokal über den Kommunikationsport die Daten "abzuzapfen" und andernorts zu übergeben. "These intermediary data transfer devices or 'milking machines' have become a popular way of solving interfacing problems" /Holloway o.J., 12/. Diese Maschinen betont Holloway sind eine typisch britische Entwicklung /vgl. ebd., 34/.

The first device, Konnect II, appeared on the market around two years ago. The manufacturer is a British company, Interset

³⁷ Ein gesondertes Problem bilden die Trennstriche, die ebenfalls herausgezogen werden müßten, ohne daß gleichzeitig Minuszeichen und Bindestriche mithinausgeworfen werden. In der Regel wird der Autor gebeten, auf Trennungen zu verzichten.

... . Konnect II uses a mini-cassette for storage of up to 64000 characters" /ebd, 34/.

Inzwischen gibt es auch andere Hersteller und Geräte, die zur Speicherung Disketten verwenden und eine größere Kapazität aufweisen.

In der Bundesrepublik wurden speziell an den Teletexdienst Erwartungen geknüpft, als Dienst für die Fremddatenübernahme im Druckbereich zu fungieren.³⁸

Als Lösung des Problems werden auch sogenannte Konverter angeboten; damit sind Computer gemeint, die speziell für die Konvertierung von Datenformaten und Datenträgern gebaut wurden.

Sie enthalten meistens Diskettenlaufwerke für Disketten unterschiedlicher Größe und einen Ausgang zur Telekommunikation. GEVA, Shaftstall, Applied Data Communication, Cromwell, Itek und Altertext gehören zu den bekannteren Herstellern von Konvertern /Vgl. Alexander u.a. 1985/. Konverter können in der Regel verschiedene Diskettenformate (Anhaltspunkt ca. 400) lesen und in Speicherformate von Satzrechnern verwandeln. Die Leistungsfähigkeit von Konvertern bemißt sich danach, wieviele Datenformate er lesen und schreiben kann; entscheidender ist noch, über welche Spannbreite unterschiedlicher Betriebssysteme sich die Konvertierleistung erstreckt.

Ein besonderes Qualitätsmerkmal eines Konverters ist die "Übernahme von Gestaltungsmerkmalen" /Knoth 1986a, 508/, d.h. die Umsetzung von Formatiercodes eines Textverarbeitungssystems in die Befehle eines Satzrechners.

Probleme machen Textverarbeitungssysteme, die die Formatierdaten in einer vom Text getrennten Datei speichern /ebd, 510/. Sie erleichtern einerseits die Fremddatenübernahme, weil der reine Text, wie bei *Word*, in ASCII-Format abgespeichert werden kann, andererseits kann die Datei mit den Formatierbefehlen nicht von einem Konverter interpretiert werden, weil die Formatiercodes darin ohne

³⁸ "Die sehr personalintensive Texterfassung kann zunehmend dem Kunden, insbesondere solchen mit größerem Textaufkommen, überlassen werden. Die Kommunikation über das ab 1981 verfügbare Teletex-System wird diesen Dialog erleichtern" /Teichmann u.a. 1982, 229/. Zur Zeit läuft eine Studie des ISI (Fraunhofer Institut für Systemtechnik und Innovationsforschung, Karlsruhe) im Auftrag der Deutschen Bundespost, die u.a. erforscht, "ob Teletex ein Markstein für die Fremddatenübernahme sein wird" /Bierhals, Steppat 1985, 28/.

direkten Zusammenhang zu den Textteilen, auf die sie sich beziehen, abgespeichert sind.

Ein weiteres Problem, das noch nicht gelöst ist, ist die Datenübernahme von Texten aus integrierten Softwareprodukten wie Lotus 1-2-3, Multiplan u.ä.

Die Konverter sind nicht billig (Anhaltspunkt 60000.- DM und mehr), so daß es sich nicht für jeden Satzbetrieb lohnt, solch einen Spezialcomputer anzuschaffen. Ein weiterer Einwand gegen ihre Anschaffung ist für einen Verlag, "that the disc you get is the one the manufacturer has not yet written a piece of software to convert" /Ferris 1983, 58/. Aus dieser Situation heraus haben sich spezielle Dienstleistungsunternehmen dieser Aufgabe gewidmet.³⁹ Eine Studie aus Großbritannien, in der ein Testdokument, auf verschiedenen Textverarbeitungssystemen erstellt, verschiedenen Druckbetrieben zur Weiterverarbeitung gegeben wurde, ergab:

"The study showed that many of the technical difficulties in phototypesetting word processor copy had been overcome. All the systems investigated as case studies were able to produce the test text as required, with minor omissions in some cases" /Holloway o.J, 1/.⁴⁰

Als Hauptlösungen für das Schnittstellenproblem werden

1. Konverter (multi-disc readers)
2. Datentransfergeräte ('milking machines') und
3. Datenfernübertragung

genannt /vgl. ebd., 15/.

Dennoch, und hier liegt der weitere Gang der Arbeit begründet, sind mit dieser Konvertierung nicht alle Probleme gelöst. Um es einfach zu sagen: **Zwar** können die Schwierigkeiten, unformatierte Texte (reinen ASCII-Zeichensatz z.B.) von Rechner zu Rechner zu portieren, als technisch gelöst angesehen werden. Verschiedene Wege wur-

³⁹ In der bereits erwähnten Teichmannstudie wird diese Spezialisierung als zukunftssträftig erachtet /vgl. Teichmann u.a. 1982, 239f/. Der Ort der Konvertierdienste ist auch aus Abbildung 1 auf Seite 8 und Abbildung 11 auf Seite 110 abzulesen.

⁴⁰ Zu berücksichtigen ist hier allerdings, daß der Text auf komplexe Textelemente wie mathematische oder chemische Formeln und Abbildungen verzichtete.

den beschrieben, die allerdings verschieden hohe Kosten verursachen.

Die Einsparnisse, die durch die Datenübernahme derart erzielt werden, beziffert S. Roth **aber** auf weniger als 20% der Satzkosten. und er führt aus:

"To achieve greater savings, both in time and money, it's necessary to go the next step in preparation - inserting codes. ... If you do all your own coding, you can expect to save between 25 and 50 percent on typesetting costs" /Roth S. 1985a, 128/.

Das Codieren in dem hier angesprochenen Sinn dient dazu, den Satzbetrieb nicht nur von der Erfassungsarbeit zu befreien, sondern auch die Eingabe der Satzbefehle zu unterstützen oder sogar zu automatisieren.

In den Ausführungen über die Konvertierer haben wir schon gesehen, daß es möglich ist, die Formatierbefehle eines Textverarbeitungssystems in Satzbefehle umzuwandeln. Wir haben aber auch gesehen, daß dieses Verfahren nicht für alle Textverarbeitungssysteme anzuwenden ist. Auch muß man sich klar machen, daß dieses Verfahren keine allgemeine Lösung ist, da sonst für jedes Textverarbeitungssystem zu jedem Setzsystem und für jedes Setzsystem zu jedem Textverarbeitungssystem Konvertiererroutinen geschrieben werden müßten.

Gerade für Verlage, die mit einer Vielzahl Autoren mit entsprechend vielfältigen Textverarbeitungssystemen zu tun haben und die womöglich auch mit verschiedenen Druckbetrieben zusammenarbeiten wollen, ist diese Situation unbefriedigend.

So haben große Verlage begonnen (z.B. Wiley, SpringerVerlag), Auszeichnungsrichtlinien zu entwickeln, d.h. Codierschemata, deren Codes die Rolle "elektronischen Markups" übernehmen und die stellvertretend für später daraus zu generierende Satzbefehle in den Text eingestreut werden. Diese Codes sind so beschaffen, daß sie geeignet für den Informationstransfer sind (d.h. nur aus Standardzeichensätzen wie z.B. ASCII aufgebaut).

Die Bemühungen, diese Codierschemata zu standardisieren, so daß Texte *mit Markup* zwischen Autoren, Verlagen und Satzbetrieben problemlos ausgetauscht werden können, werden in den nächsten Kapiteln ausführlich behandelt. An dieser Stelle soll noch ergänzt werden, daß diese Codierschemata auch verwendet werden können, um in Elektronische Manuskripte mathematische und chemische Formeln und Tabellenmaterial in codierter Form zu integrieren und somit auszutauschbar zu machen. Daß diese Codes auch eine Rolle bei der Entwicklung elektronischer Lieferformen spielen, wird ebenfalls später deutlich gemacht.

3.1.3 Die Autor - Verlagsbeziehung

Das Elektronische Manuskript kommt zum Verlag in einer Form, in der es nicht bleiben kann. Gerade wenn man die Rolle des Verlags in der "Beratung und Kontrolle" des Autors /Fabian 1983, 293/ sieht, ist evident, daß ein Verlag ein Manuskript *nicht* unverändert übernehmen kann.⁴¹ Es fragt sich allerdings, ob schon in dieser frühen Stufe das Manuskript elektronisch angeliefert werden sollte.

"Submitting the initial manuscript electronically is of minimal value to the overall publishing cycle *unless* a means to incorporate editorial changes, revisions, and markup can also be found electronically before the manuscript, in electronic form, is typeset in page form" /Myers 1984, 10/.

Das Problem ist, daß die Veränderungen und das Markup, das im Verlag angebracht wird, sich dem Autor mitteilen muß, ohne die Identität der Vorlage zu zerstören. Ähnlich ist auch das Problem gelagert, wenn das Manuskript durch mehrere Hände (über mehrere Terminals) gehen soll, also mehrere Personen Änderungen einzubringen haben.

Tatsächlich gibt es inzwischen Software, die erlaubt, Bemerkungen in einem Text oder sogar "auf dem Rand" anzubringen, ohne die Identität der Vorlage zu zerstören und mit der Möglichkeit, Bemerkungen und Text getrennt oder gemeinsam auszugeben.⁴² Der Nachteil ist bislang, daß diese Software teuer ist und voraussetzt, daß Autor und Verlag dieselbe Hard- und Software besitzen, was wohl eher die Ausnahme sein wird.

⁴¹ Es herrscht eine recht weit verbreitete Ablehnung gegen die Praxis, "reproreife" Typoskripte zu veröffentlichen. Hier wird die Selektionsfunktion und Qualitätskontrollfunktion des Verlags weitgehend umgangen. Daneben krankt das vervielfältigte Manuskript an mangelnder Lesbarkeit und gerade bei Sammelbänden an einem Wildwuchs verschiedener Gestaltungsprinzipien und -stile. Der große Vorteil dieses "quick and dirty" Verfahrens ist seine Schnelligkeit. Vom Elektronischen Publizieren wird erwartet, daß es Schnelligkeit *und* Qualität wieder zusammenbringt. Vgl. statt anderer /Fabian 1983, 293-295/.

⁴² S. Roth nennt den "Wang word processor", das "Publisher Six System" von Computek und "Bookware" von Keytext als Beispiele /Roth S. 1985a, 114/.

In der Regel wird diese Stufe der Manuskriptbearbeitung noch traditionell erledigt, d.h. auf Papier. Damit wird nicht nur das Problem des elektronischen Markup umgangen, sondern auch die Arbeit am Manuskript erleichtert, da ein Text auf Papier übersichtlicher und leichter zu verfolgen ist als am Bildschirm. Zwei Formen der Autor - Verlagsbeziehung leiten sich daraus her.

Im ersten Verfahren nimmt der Verlag Veränderungen auf Papier vor und der Autor arbeitet sie in sein Elektronisches Manuskript ein und liefert schließlich ein fehlerfreies Elektronisches Manuskript ab.

Die Kontrolle über den Werdegang des Manuskripts bleibt so beim Autor. Das Konvertierproblem verschiebt sich auf den Übergabemoment des Elektronischen Manuskripts. Das Recht des *Imprimatur* wird aufgeweicht. Da der Autor nicht mehr wie früher in die Korrektur der Satzfarben eingebunden ist, bekommt er sie auch nicht mehr zur letzten Kontrolle zu Gesicht. Für ihn ist die Kontrolle über sein Manuskript mit der Abgabe der Diskette abgeschlossen - Sonderregelungen ausgenommen.

Ohne großes Aufheben hat der Autor außerdem sukzessive alle Korrekturen, die vom Verlag vorgeschlagen wurden, selbst in seinem Text vorgenommen.

Früher, damit komme ich zum zweiten Verfahren, nahm die Setzerei die Änderungen am Manuskript zusammen mit der Neuerfassung vor. Dieser Weg ist auch gangbar, wenn der Autor seine Diskette abgeliefert und der Verlag dem Autor seine Änderungsvorschläge auf einem Ausdruck anmerkt, der Autor sie akzeptiert oder verändert zurückschickt und der Verlag das Einfügen der Korrekturen übernimmt bzw. an den Satzbetrieb delegiert. In diesem Fall behält der Verlag die Kontrolle über den Werdegang des Manuskripts und dem Autor werden keine Korrekturen übertragen. Andererseits wird jeder Autor Wert darauf legen, sein Elektronisches Manuskript auf dem neuesten Stand zu halten und darum parallel zum Verlag Veränderungen selbst vornehmen.

In allen drei genannten Varianten der interaktiven Manuskriptbearbeitung zwischen Autor und Verlag taucht früher oder später das Konvertierproblem auf, wenn man von dem seltenen Fall absieht, daß Autor und Verlag bzw. Satzbetrieb kompatible Maschinerie einsetzen. In allen drei Fällen erspart der Autor dem Verlag die Neuerfassung (zumindest teilweise) und sich die Korrektur der Satzfarben.⁴³

⁴³ Hier ließe sich über eine Entlohnung der Autoren für ihre geleistete Erfassungsarbeit nachdenken und darüber, wer für die Konvertierung der Daten aufkommen muß.

Die Autor-Verlagsbeziehung ändert sich entscheidender, wenn die Autoren in den Herstellungsprozeß miteinbezogen werden. Das ist zwar auch der Fall, wenn sie ihre maschinenlesbaren Manuskripte abgeben, aber es bleibt ihnen ebenso äußerlich, als wenn ein Verlag hergeht und ein maschinenschriftliches Manuskript über einen Blattleser einliest.

Der tatsächliche Einbezug in den Herstellungsprozeß wird über die Auszeichnungsrichtlinien ventiliert. Wenn der Autor jetzt Codes in den Text einstreut, geschieht das (bewußt) in Hinblick auf die Erleichterung der späteren Weiterverarbeitung und ist im Idealfall der Ansatzpunkt für die Automatisierung der Satzherstellung. Es ist zu früh, hier über die Lastenverteilung oder Lastenumverteilung im Publikationsprozeß zu sprechen, da es eine Reihe von Variablen gibt, die zu berücksichtigen sind, nämlich in welcher Weise das Einstreuen des Codes arbeitsteilig zwischen Autor, Verlag und Satzbetrieb vorgenommen wird, welche Softwareunterstützung für die Eingabe angeboten wird u.a.. Die Frage wird in " 4. Elektronisches Publizieren mit der SGML" auf Seite 66 besonders "Anwendungen der SGML" auf Seite 86 und " 5. Schlußbetrachtungen" auf Seite 102 weiterverfolgt.

3.1.4 Das Verhältnis Satz - Verlag

Das Verhältnis Satz - Verlag wird, wie erwähnt, direkt von den Elektronischen Manuskripten betroffen. Auf Fremddatenübernahme vorbereitet zu sein und z.T. Spezialisierung auf Konvertierdienste waren die angeratenen Empfehlungen auf eine Zunahme von Fremddaten und damit Abnahme der Erfassungstätigkeit in den Satzbetrieben.

Um den Vorgang in seiner Dimension richtig einschätzen zu können, einige Sätze und Zahlen zur Bedeutung der Verlage für den Druckbereich, zu dem neben den eigentlichen Druckereien, die Reprofirmen und Satzbetriebe zählen - Mischformen sind üblich.

Kaufmännisch betrachtet ist der Verleger Auftraggeber der Druckbetriebe, oder andersherum der Verleger Kunde des Druckers. Was bedeuten also die Verlagsprodukte für den Druckbereich? Von einem Umsatz von ca. 18.838 Millionen DM (für 1979) entfielen auf Bücher und Broschüren ca. 1.5 Millionen, auf Zeitschriften ca. 3.0 Millionen und auf Zeitungen ca. 3.5 Millionen /vgl. Teichmann u.a. 1982, 58 Quelle dort: Statistisches Bundesamt/.

Weniger als die Hälfte des Umsatzes des Druckbereichs wird somit durch Verlagsprodukte erzielt. Daten einer Ifo-Befragung von 1975 kommen zu einem Anteil von 35,7 % die der Kunde "Verlag" für den

Umsatz des Druckbereichs erbringt /vgl. Teichmann u.a, 1982, 51 (Quelle dort: Ifo-Sonderbefragung 1975)/.

Dabei wird ganz deutlich, daß die Verlage als Kunden immer wichtiger werden, je größer der Druckbetrieb:

Daten aus der Ifo-Sonderbefragung 1975 im Auszug
/vgl. Teichmann u.a. 1982, 51f/ belegen das.

KUNDE	BETRIEBSGRÖSSE			
	bis 49	50-200	200-499	500 u.mehr
Industrie	51,1%	46,6%	40,0%	16,9%
Verlage	15,6%	20,7%	36,0%	46,7%

Umsatzanteile von Druckbetrieben unterschiedlicher Größe bezogen auf die Auftraggeber Industrie und Verlag für 1975⁴⁴

Diese Zahlen werden bestätigt, wenn man sich die Produktionsschwerpunkte in verschiedenen Betriebsgrößenklassen ansieht /vgl. Teichmann u.a. 1982, 55/.

Dem Trend, Einbußen durch die Rückverlagerung der Erfassungstätigkeit an den Ort der Entstehung der Texte zu erleiden, wurde die Besinnung auf die spezifischen Qualitäten der Satzbetriebe entgegengehalten (vgl. Fußnote²⁹). Die Hauptgefahr für den Satzbereich ist darin zu sehen, daß man beginnt, auf diese *spezifischen* Leistungen zu verzichten, die Fremddaten also gar nicht mehr bei den Satzbetrieben ankommen.

Diese Gefahr kommt aus dem Bereich des "Inhouse-Publishing" und dem Bürobereich. Selbstverständlich geht es hier nicht nur um Manuskripte, sondern auch um Formulare etc.. Andererseits ist aber auch davon auszugehen, daß Autoren nicht nur zu Hause an ihrem PC sitzen, sondern in Organisationen eingebunden sind, die Schreibbüros und Hausdruckereien zur Verfügung haben.

⁴⁴ Es darf aber nicht ganz außer acht gelassen werden, daß die Verlage als Kunden, nicht nur typische Verlagsprodukte in Auftrag geben.

"Der weitere Aufbau von Hausdruckereien bei großen Firmen/Institutionen bereitet ... Sorgen" /Teichmann u.a. 1982, 24/

und

"Die technischen Möglichkeiten, die Betreibern von Bürotextsystemen demnächst durch zusätzliche Software (Gestaltungsprogramme) und Hardware (Belichter) gestatten, eigenen 'Satz' zu erzeugen, müssen als die Konkurrenz gesehen werden, die sie in der Tat für den Satzbetrieb darstellen" /ebd., 225/.

Diese Zitate drücken ernstzunehmende Entwicklungslinien, die jenseits des Druckbereichs verlaufen, aus; sie sind für den Satzbereich viel drückender als eventuelle Entwicklungen im Bereich elektronischer Lieferformen von Publikationen.

Die technischen Voraussetzungen dieser Entwicklung sind ambivalent. Wenn zunehmend Satz auf PCs erzeugt werden kann, bedeutet das doch einerseits, daß auch die Satzbetriebe günstigere Geräte einsetzen können, um Satz zu erzeugen. Andererseits bedeutet die Entwicklung aber eben auch, daß Satzerstellung nicht mehr an den Satzbetrieb gebunden ist.

Die Hersteller von Setzsystemen jedenfalls haben begonnen, Mikrocomputer allmählich in ihre Systeme einzubeziehen, "angefangen bei *AM International*, über *Berthold*, *Compugraphic*, *Linotype*, *Scangraphic*, *Siemens bis hin zu Xyvision*" /Fritz 1986b, 806/. Das heißt, im Satzbereich wird zunehmend mit den Geräten gearbeitet, über die auch der Autor verfügt.

PC-Setzsysteme lassen sich danach unterscheiden, wieweit sie Satzarbeiten übernehmen: Ausschließen, Trennen, Seiten umbrechen und gestalten und gegebenenfalls auch Text-Graphik-Integration. Fritz unterscheidet nach Funktionsumfang "redaktions-, produktions- und gestaltungsorientierte PC-Setzsysteme" /Fritz 1986a, w78/, einer Einteilung eines amerikanischen Setzerverbandes folgend.

Ready Set Go, MacPublisher, PC-TeX, PagePlanner, Buchmaschine, Magna, Superpage, Do-it, Maxx sind einige solcher Systeme, die Prechsel bespricht /vgl. Prechsel 1986/. Auch in den genannten Artikeln von Fritz werden einige PC-Setzsysteme besprochen. Für den amerikanischen und britischen Markt gibt es nach Fritz ca. 30 PC-Satzprogramme /vgl. Fritz 1986b, 804/.

Diese Systeme lassen zwar die Grenze zwischen Autoren- und Verlagssystemen fließend werden und erlauben dem Autor, Satz zu erstellen, aber spätestens bei der Frage nach der Fotosatzbelichtung zeigt sich die Herstellerabhängigkeit der Setzsysteme, so daß ein

Autor seinen Satz nicht zur Belichtung und zum Druck dorthin geben kann, wohin er will, sondern dorthin, wo Geräte desselben Herstellers zu finden sind. Die "Buchmaschine" gibt immerhin an vier Fotosatzbelichter aus (Linotron 202, Lasercomp, Digiset, CRTronic 150). Zur Zeit stellen die PC-Setzsysteme aber auch noch keine vollständige Alternative für die Satzbetriebe dar, weil sie "jeweils nur einen Ausschnitt des gesamten Satzspektrums abdecken"/Fritz 1986b, 804 /.

Die Lasertechnologie ist auf dieselbe Weise ambivalent wie der Mikrocomputer als Satzrechner. Sie kann im Satzbetrieb, in Laserdruckern und Belichtern realisiert, beispielsweise Kosten einsparen, wenn für die Korrekturabzüge der Laserdrucker und erst für die entgeltliche Druckvorlage der Belichter eingesetzt wird; oder kleinere Auflagen können über den Laserdrucker ausgegeben werden und somit für den Auftraggeber attraktiver als teurere belichtete Druckerzeugnisse sein. Der Laserdrucker als ein Gerät, das auch an Mikrocomputer angeschlossen werden kann, ist aber ebenfalls nicht dem Satzbetrieb vorbehalten.

Dennoch gilt nach wie vor, daß hochwertiger Satz und entsprechend hochwertige Druckausgabe die Kenntnis des Fachmanns und teurere Ausrüstung erfordern. Diese Chance kann der Satzbereich aber nur wahren, wenn er dreifach für Fremddaten offen ist - für unformatierte Daten und "generisch codierte" Daten wie bereits angesprochen (und im Lauf der Arbeit weiter thematisiert), und für bereits auf PCs gesetzte Daten (!), die nun auf einer höherwertigen Ausgabeinheit ausgegeben werden sollen.

Dem steht allerdings die Inkompatibilität von PC-Satz und Fotosatzanlage entgegen.

Auch hier gibt es allerdings eine interessante Entwicklung, die zum Abschluß dieses Abschnitts kurz angesprochen werden soll.

Ihre technische Basis ist die Lasertechnologie, integriert in Ausgabeinheiten, *Scanbild-Ausgabeinheiten*, "die aufgrund ihrer Technologie bei der Belichtung die gesamte Breite des verwendeten Materials bestreichen. Das Material können in einem Laserdrucker Normalpapier, in einem Laserbelichter ein Film oder in einem mit UV-Licht arbeitenden Flachbettrecorder eine Offsetplatte sein."/ Schmid 1986, w2/

In die Ausgabeinheit eingebaut ist ein sog. *Raster- Image- Prozessor*, der Scanbilder generiert.

"Die Aufgabe eines RIP besteht darin, eingehende Daten so aufzubereiten, daß der angeschlossenen Scanbild-Ausgabein-

heit komplette, in Scanlinien zerlegte Daten übergeben werden können"/Schmid 1986 w2f/.⁴⁵

Eine Scanbildausgabeeinheit setzt fix und fertig umbrochene Seiten voraus. Die vom Setzsystem fertig umbrochenen Seiten mit allen Informationen über Schriften und Graphiken werden in Bildpunkte umgesetzt. Der RIP versteht eine bestimmte Eingabesprache, die in der Regel herstellerabhängig zwischen Satz und RIP vermittelt (CORA von Linotype gilt als weitverbreitet).

An dieser Stelle setzt POSTSCRIPT als Seitenbeschreibungssprache an:

"It provides a tool that a composition program can use to describe pages. So a CAD/CAM program can use output POSTSCRIPT to do CAD/CAM drawings. A text layout program like TECor Scribe or Interleaf can build descriptions in POSTSCRIPT to describe how the page is to be put together. So it is an intermediate bridge between application programs and the final marking engine" /Warnock 1984, 4-17/.

Für verschiedene Anwendungsprogramme wird eine Seitenbeschreibung erzeugt, die von vielen Scanbild-Ausgabeeinheiten mit Rasterimageprozessoren verstanden werden kann. Ein Satzprogramm wird in die Lage versetzt, seinen Satz in "POSTSCRIPT" auszugeben, und ein RIP ist in der Lage, diesen Output als Input zu verarbeiten. Statt der Relation *eín* Setzsystem *eín* Ausgabegerät, stehen nun *x* Anwenderprogramme *y* Ausgabegeräten gegenüber, ohne inkompatibel zu sein.

Wenn sich POSTSCRIPT⁴⁶ als Industriestandard durchsetzen sollte, bedeutete das, daß a) von einem PC-Satzprogramm aus Laserdrucker und/oder Belichter angesteuert werden können und daß b) es die Laserdrucker oder Belichter verschiedener Hersteller sein können. Beliebtes Beispiel für eine solche Konfiguration ist der Macintosh von Apple, dessen Satzprogramm 'PageMaker' in der Lage ist, "POSTSCRIPT" zu erzeugen, das von dem Apple-LaserWriter und Belichtern von Linotype verarbeitet werden kann - mit der den Geräten eigenen Auflösung.

⁴⁵ Hier muß das Thema "RIP" in zwei Sätzen abgehandelt werden. Für mehr Information sollte zunächst der Artikel von Schmid herangezogen werden, wegen seiner (wohltuenden) Klarheit.

⁴⁶ Vgl. weiter zu POSTSCRIPT /Bechtolsheim 1986 a und b/. Siehe auch /Scheller, Appelt 1986, 9-11./

Es steht zur Diskussion, wie die künftige Aufgabenteilung in konkreten Fällen zwischen den Orten, wo die Daten anfallen, und den Satzbetrieben aussehen wird.

Das Kompatibilitätsproblem, das sollte auch mit dem Schlenker zu POSTSCRIPT gezeigt werden, ist keineswegs trivial, weil es nicht nur um den reinen Datenaustausch geht, sondern auch darum, daß die Daten stufenweise weiterverarbeitet werden müssen vom Manuskript bis zum physischen Output. Dem Text werden die Formatierinformationen und schließlich auch die Informationen zur Steuerung der Ausgabeeinheiten hinzugefügt.

Im nächsten Abschnitt wird ein Modell dargestellt, das einen systematischen Zugang zu den Fragen des Datenaustausches und den Stufen des Dokumenterstellungsprozesses bietet. Mit der Berücksichtigung unterschiedlicher Textelemente wird die Komplexität modelliert, die integrierte elektronische Publiziersysteme in der Praxis zu bewältigen haben.

3.2 EIN ALLGEMEINES MODELL DER DOKUMENTERSTELLUNG

Der Prozeß der Herstellung von Dokumenten, egal ob für den Bürobereich oder die Öffentlichkeit bestimmt, läßt sich in das ISO-Referenzmodell integrieren.⁴⁷ Eine klare Beschreibung der Zielsetzung findet sich bei Meißner:

"Ziel der Entwicklung und Standardisierung des Referenzmodells für die Kommunikation Offener Systeme war die Schaffung einer gemeinsamen Basis bzw. eines Architekturmodells, in dem die verschiedenen Aspekte der Kommunikation zwischen Rechnersystemen im Detail beschrieben und standardisiert werden können. ... Unter dem Begriff **Offenes System** - eine bewußte Abstraktion - wird ein System, bestehend aus einem oder mehreren Computern, der dazugehörigen Software, Peripherie, den Terminals, Operator etc. verstanden, das "offen" ist für die Kommunikation mit beliebigen anderen (offenen) Systemen. Offen ist die Kommunikation dann, wenn sowohl die logischen als auch die physikalischen Verbindungen losgelöst von Implementierungsaspekten sowie spezifischen Eigenschaften der

⁴⁷ Das Referenzmodell ist der Internationale Standard 7498 der International Standard Organisation (ISO). Siehe /ISO 7498 1984/. Manchmal wird das Modell auch ISO/OSI-Modell genannt. Beide Bezeichnungen ergeben sich aus dem Titel der Norm: "Open Systems Interconnection - Basic Reference Model"

Systeme definiert sind und die Kommunikation nach einem Standard erfolgt, der allen Systemherstellern zugänglich (z.B. nicht durch Patente geschützt) ist." /Meißner 1985, S.55-88/

Eine plastische Darstellung der sieben Schichten des Modells und eine Übersicht über die Normungsaktivitäten der ISO "über alle Schichten" zeigt die Graphik auf der folgenden Seite, die /Blo-meyer-Bartenstein, Both, 46/ entnommen ist.

Eine Standardisierung der Dokumenterstellung wäre auf der Schicht 7, der *Anwendungsschicht* anzusiedeln. Diese Schicht ist bislang kaum standardisiert. Nach Goergen u.a. "können bestenfalls einzelne typische Anwendungen vereinheitlicht werden... . So wird gerade diese Schicht in eine Mehrzahl anwendungsspezifischer Teilschichten oder Gruppen von Funktionen zerfallen" /Goergen u.a. 1985, 45/.

Die Modellbildung, die im folgenden kurz skizziert wird, ist *nicht* Teil der *ISO/OSI*-Normung, sondern ein Versuch, den Dokumenterstellungsprozeß *nach Prinzipien* des Schichtenmodells - analog - zu entwerfen. Ich folge hier im Ansatz und z.T. in der Darstellung Petersen /Petersen 1985/. Der Bezug zu dem aktuellen Normungsstand besteht insofern, als der *Austausch* von Dokumenten in offenen Systemen selbst Gegenstand der internationalen Normung ist (vgl. Abbildung 5 auf Seite 49). Die wichtigen firmenunabhängigen Austauschformate für Dokumente, z.B. ODIF (Office Document Interchange Format) und SDIF (SGML Document Interchange Format) sind Mittler zwischen der "OSI-Normung" und dem normierten Dokumentmodell ODA (Office Document Architecture) bzw. der SGML (Standard Generalized Markup Language). Zweck der Dokumentarchitekturen bzw. der SGML ist es, auf Ebene der *Datenstrukturen* Kompatibilität zu erreichen.⁴⁸ Es sei kurz erwähnt, da im folgenden vorausgesetzt, daß die Entwicklung einer ISO-Norm einem zeitaufwendigen Procedere nachkommt, oder wie es Joan Smith sagt: "The making of ISO standards is a democratic process, and a lengthy one"/Smith 1986 c, 2/. Ein Normprojekt beginnt bei den "Working Drafts" (WD), wird dann zum "Draft Proposal" (DP), danach zum "Draft International Standard" (DIS) und schließlich zum "International Standard" (IS).

⁴⁸ Der Standard ECMA-101 Office Document Architecture der European Computer Manufacturer Association wird in der ISO als Draft Proposal 8613 noch weiterverfolgt und ausgebaut. Der Normvorschlag ist bei der ISO mehrteilig. Teil zwei enthält die "Office Document Architecture", Teil 4 das "Office Document Interchange Format". Insgesamt enthält der Normentwurf neun Teile. Auf die ISO/DIS 8879 (SGML) komme ich gleich in "Die ISO/DIS 8879 (SGML)" auf Seite 66 zu sprechen.

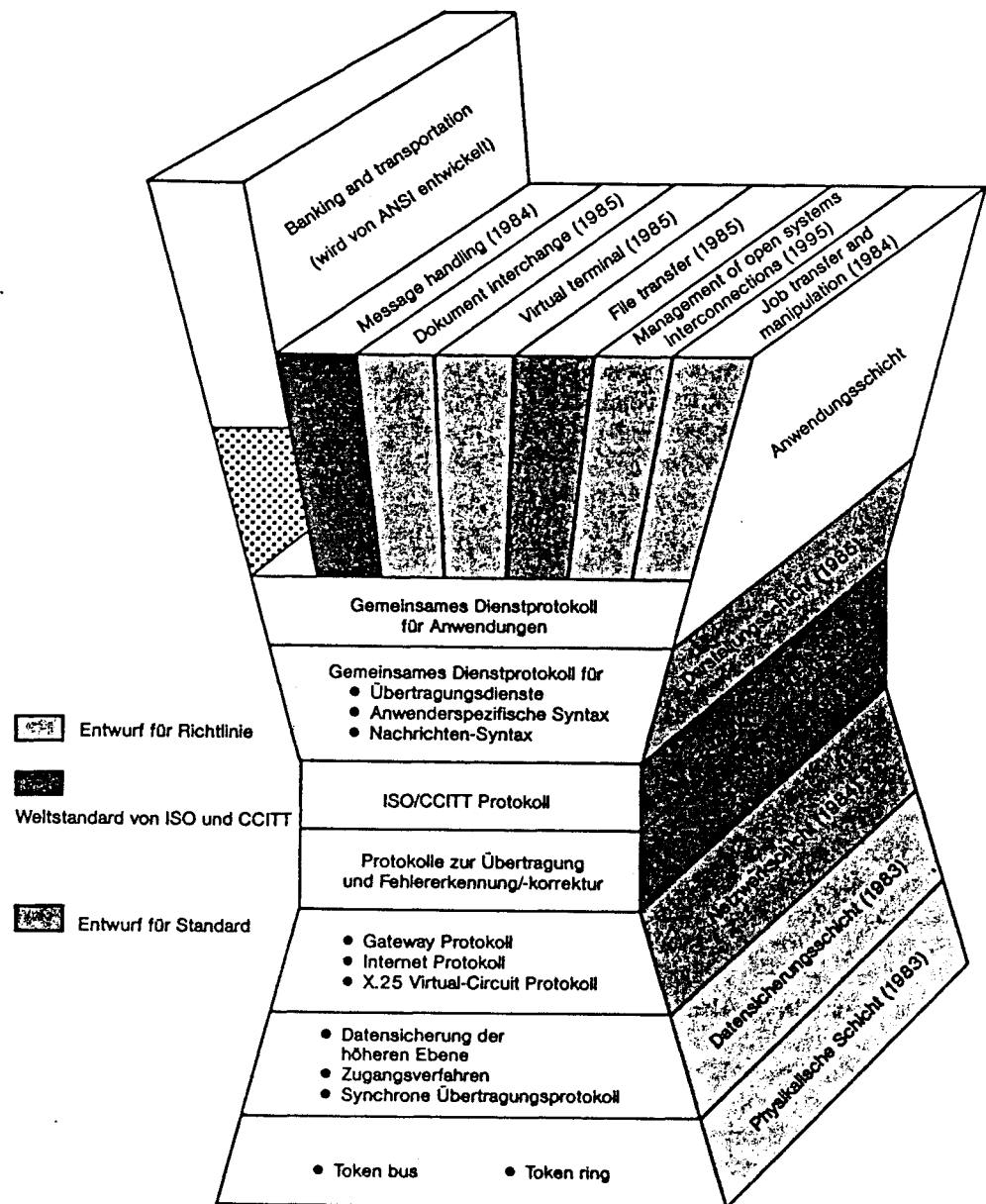


Abbildung 5. Stand der ISO/OSI Normung.

Ein Modell des Dokumenterstellungsprozesses macht Sinn, wenn es sich in den gesteckten Rahmen von Dokumentarchitektur und Dokumentaustausch einpaßt - quasi als Binnendifferenzierung. Ein Modell, daß den Dokumenterstellungsprozeß als Differenzierung der Anwendungsschicht in weitere Schichten konzipiert, kann dem entsprechen.⁴⁹

Es müssen sich dann Teilprozesse identifizieren lassen, die, mit bestimmten notwendigen Funktionen ausgestattet, im Sinne des Schichtenmodells miteinander kommunizieren. Wieweit einzelne Funktionen von Menschen oder Maschinen erfüllt werden, hängt auch hier, wie im Modell generell, von konkreten Implementationen ab. Zwischen den Teilprozessen, die als Schichten aufzufassen sind, müssen definierte Übertragungsvorgänge derart stattfinden, daß die nächst höhere Schicht auf den bereits erbrachten Leistungen aufsetzen kann. In der Terminologie des Modells wird von Schichtenprotokollen oder Protokollen gesprochen als einer "Menge von Regeln und Datenformaten, mit deren Hilfe zwei Partner-Instanzen über eine Verbindung gemeinsame Funktionen erfüllen, d.h. innerhalb einer Schicht den geforderten Dienst realisieren." /Goergen u.a. 1985, 52/

In unserem Zusammenhang ist es angebracht, die Kommunikation der Anwendungsinstanzen nicht nur zweiseitig anzusehen, sondern als Kooperation verschiedener Anwendungsinstanzen in unterschiedlichen Systemen. Von solch verteilter Anwendung auszugehen, entspricht dem vielstufigen Dokumenterstellungs- bzw. Publikationsprozeß. Als Sonderfall wäre ein geschlossenes System wie die Schreibmaschine zu betrachten.

Eine Abbildung des Schichtenmodells, das die Dokumenterstellung berücksichtigt, folgt umseitig.

Das "Urmodell" der Abbildung entspricht /ISO 7498 1984, 19/. Eine Abbildung bei /Petersen 1985, 179/, von der ich ausgegangen bin, konkretisiert das Schema für ein zweipoliges Modell von Autor/Dokument - Autor/Dokument. Hier (Abbildung 6 auf Seite 51) werden drei Anwender Autor, Setzer, Drucker illustrierend benannt. Die Bezeichnungen sollen aber nicht dazu verleiten, die einzelnen Stufen mit traditionellen Berufen zu belegen oder anders, einzelne Berufe an bestimmte Funktionen zu binden. Aus der Abbildung wird überdies deutlich, daß die Dokumenterstellung sich durch die verschiedenen Darstellungsformen differenziert und deren Integration als zusätzliches Problem aufwirft.

⁴⁹ Die Möglichkeit, Schichten zu differenzieren, "sublayers" zu bilden, ist durchaus auch im OSI-Modell vorgesehen. Vgl. /ISO 7498 1984, 6/

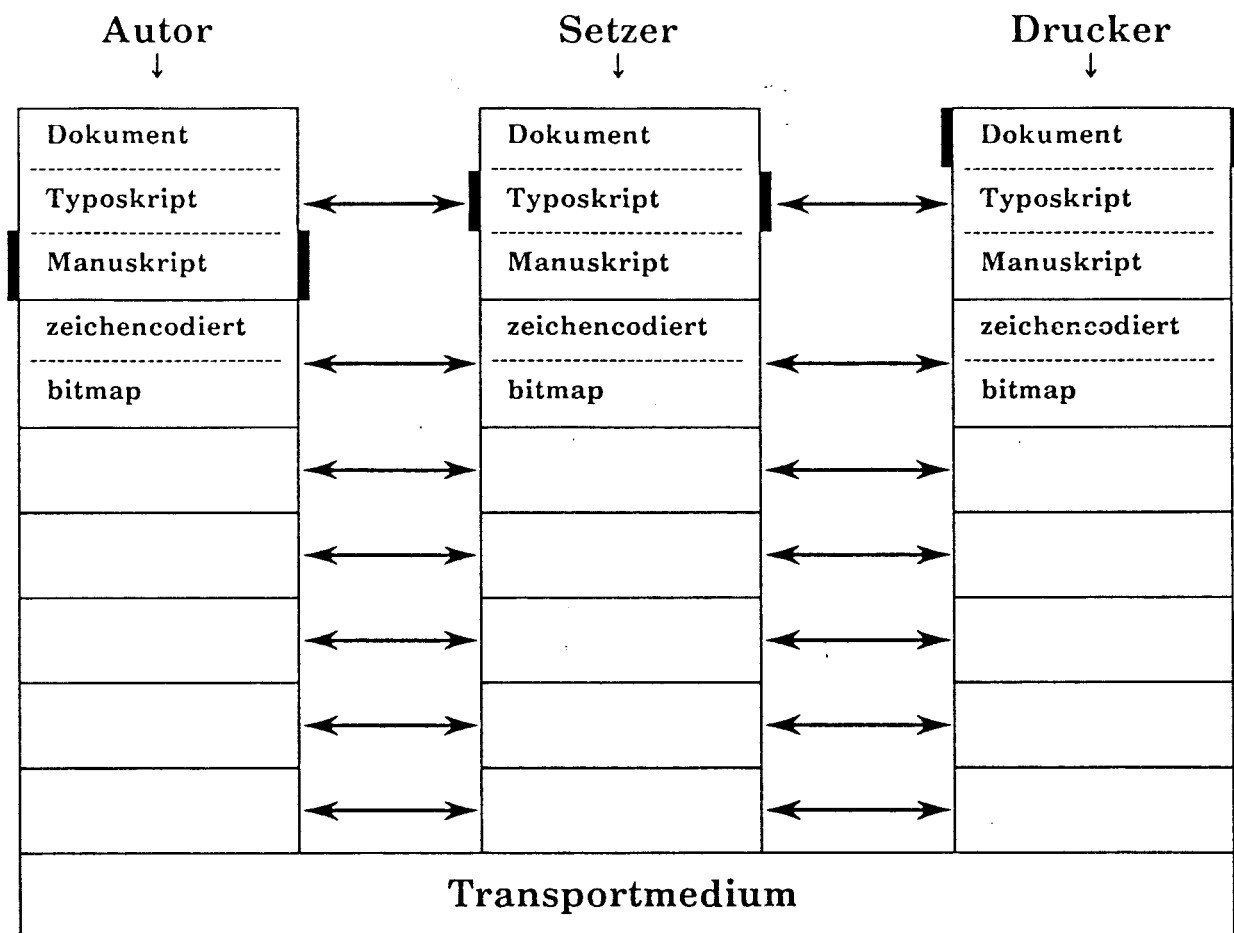


Abbildung 6. Schichtenmodell der Dokumenterstellung

Für den Bereich der Dokumenterstellung lassen sich drei Herstellungsstufen mit je spezifischen Funktionen unterscheiden:

- Manuskript
- Typoskript
- Dokument

Am Anfang steht die Eingabe oder Erfassung (Mensch-Maschine-Schnittstelle), die von ihrer Grundfunktion her die Beschreibung eines Inhalts bedeutet. Diesen Inhalt - die message, wenn man so will - identisch gesetzt, kann die Beschreibung des Inhalts noch

sprachlich gestaltet, auf grammatikalische und orthographische Fehler durchgesehen werden. Das Manuskript enthält danach die vollständige, sprachlich gestaltete Information in syntaktisch und orthographisch richtiger Form: "alle Informationen, die der Autor vermitteln möchte" /Petersen 1985, 164/. Ob durch Arbeitsteilung von Autor und Sekretariat die Funktionen "Produktion von Information" und "Erfassung" getrennt sind, ist praktisch relevant, im Modell jedoch nicht.

Enthalten im Manuskript ist seine *logische Struktur*, die i.d.R. *hierarchisch, sequentiell* /vgl. Horak 1985, 184f/ aufgebaut sein wird und graphisch als Baum dargestellt werden kann. Titelblatt, Inhaltsverzeichnis, Vorwort, Kapitel, Bibliographie wären beispielsweise logische Elemente, die auf *einer* Hierarchiestufe liegen, deren Vorkommen aber sequentiell erfolgt. Unterkapitel, Absatz, Satz, Wort, Buchstabe wäre eine mögliche Hierarchie von Elementen innerhalb eines übergeordneten Elements Kapitel.⁵⁰

Um auf den Punkt zu kommen: "Ein Manuskript an sich (nicht dessen Darstellung!) enthält zunächst keine Gestaltungsinformation, besitzt also keine physikalische (oder typografische) Struktur." /Petersen 1985, 165/ Das Manuskript ist nach diesem Verständnis weder auf eine bestimmte Gestaltung noch auf ein bestimmtes Ausgabegerät mit entsprechenden Datenträgerqualitäten bezogen. Die gestaltungs- und ausgabegerätbezogenen Informationen werden erst in den weiteren Herstellungsschritten hinzugefügt.

Die logische Struktur des Dokuments herauszuheben und die Layout-Struktur auszuklammern ist für den "Schreibmaschinegewöhnten" zumindest eine Umstellung. Beim Schreibmaschinenschreiben bedarf es keiner analytischen Trennungen der Dokumenterstellungsschritte, weil Manuskript, Typoskript und physikalische Ausgabe des Dokuments in *einem* Herstellungsprozeß an *einer* Maschine integriert sind.

Um ein Manuskript zu erstellen, müssen lediglich Eingabe-, Korrektur-, Speicher-, Ausgabe- und Übertragungsfunktionen für die Elemente, aus denen ein Manuskript besteht, bereitgestellt werden. Ein Manuskript ist auch auf dieser Stufe selbstverständlich speicher- und darstellbar, wie die genannten Funktionen schon besagen; *nur*, diese Darstellungs- und Speicherformen haben in dem Konzept vom Manuskript keinen Platz, dürfen - auf konkrete Systeme bezogen - die weiteren Prozesse nicht prejudizieren.

⁵⁰ Auf logische Strukturen komme ich weiter unten zu sprechen. Ein Strukturbaum findet sich in Abbildung 8 auf Seite 77

In der zweiten Herstellungsphase bildet das fertige Manuskript - Informationen über Inhalt und logische Struktur - die Eingabe, die zum Typoskript weiterverarbeitet wird. Die Informationen, die hinzukommen, beziehen sich auf die Gestaltung des Manuskripts. Konkret illustrierend sind darunter die Aufgaben von Setzern und/oder Satzrechnern zu verstehen: Schriftwahl, Seitenumbruch, Seitenzählung, Fußnotenberechnung etc..⁵¹

Im dritten Teilprozeß erst wird das Typoskript für eine bestimmte physische Ausgabe aufbereitet; die Informationen des Typoskripts müssen auf dieser Stufe für verschiedene Ausgabegeräte übersetzt werden. Als Ausgabegeräte kommen z.B. verschiedenste Drucker, Bildschirme, Bildplatten, Belichter, Magnetbänder oder Disketten in Frage.

Der Herstellungsablauf in einem Dokumenterstellungssystem nach diesem Modell besteht also aus drei Zwischenprodukten (Speicherformat des Manuskripts, Speicherformat des Typoskripts und Speicherformat für die Ausgabe) und zwei internen oder offenen Übergabestellen. Der "Pfiff" an dem Modell liegt nun darin, daß die Übergabe durch Protokolle geregelt und so Kompatibilität zwischen den Teilsystemen erreicht werden soll. Beliebige Systeme zur Erfassung, Weiterverarbeitung und Ausgabe können so idealiter zusammengebracht werden, wenn sie die entsprechenden Protokolle einhalten bzw. Schnittstellen besitzen. Von der Softwareseite werden die verschiedenen Formate durch Textverarbeitungssysteme, Formatierer und Gerätetreiber erzeugt, wenn man einmal zur Veranschaulichung den Stufen der Herstellung gängige Programmtypen zuordnet. Das Speicherformat der Textverarbeitung muß von dem Formatierer interpretiert werden können, der formatierte Text muß von dem Gerätetreiber interpretiert und umgesetzt werden können.

Mit der folgenden Abbildung soll das Gesagte nochmal "visualisiert" werden. Darüber hinaus soll aber bereits verdeutlicht werden, daß sich das Modell erheblich dadurch verkompliziert, daß unterschiedlichste Textelemente zu berücksichtigen sind.

Obwohl komplexe Textelemente und unterschiedliche Darstellungsformen in dem Modell keine ausdrückliche Rolle spielen, sind sie immer mitzudenken, wenn man das Modell konkretisieren oder praktisch implementieren wollte. Das gilt in bezug auf die Ein- und Ausgabe als auch für die spezifische Bearbeitung der verschiedenen Textelemente auf jeder Stufe. Komplexe Textelemente und unterschiedli-

⁵¹ Wenn es gelänge, an der logischen Struktur ansetzend, diese auf eine Layout-Struktur abzubilden, wäre der Typoskripterstellungsprozeß automatisierbar.

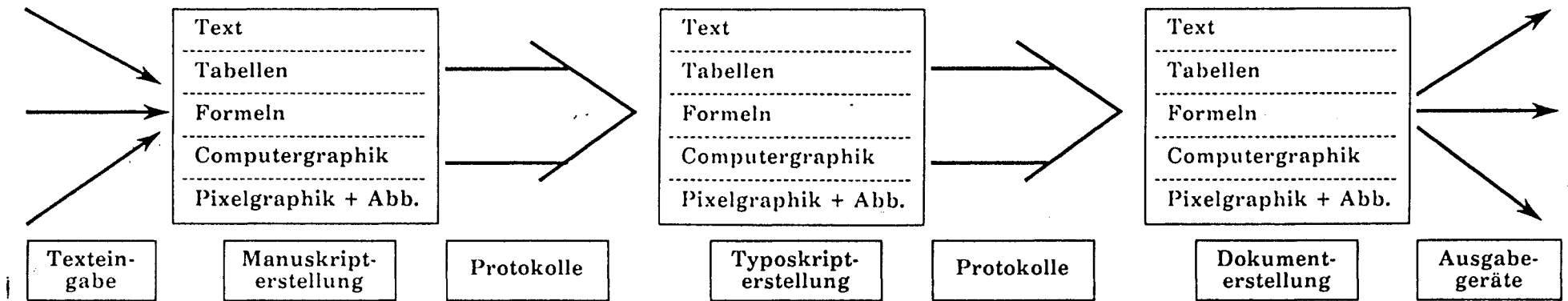


Abb.: Ablaufmodell der Dokumenterstellung für zeichen- und bitorientierte Textelemente

che Präsentationsformen multiplizieren die Probleme der Eingabe, der Bearbeitung, der Übertragung und der Ausgabe. Die Forderung, Textelemente nicht separat, sondern integriert zu bearbeiten, bringt weitere Probleme mit sich.

Ich will die verschiedenen Textelemente, die vorrangig in Betracht kommen, nochmals aufführen.⁵² Grundsätzlich bietet sich die Unterscheidung von zeichenorientierten und bitorientierten Speicherformen an. Es ist aber zu beachten, daß Zeichen auch als "bitmap" und Graphiken, Formeln und Tabellen auch zeichencodiert gespeichert werden können.

- Fließtext (lineare Anordnung) von Zeichen. Standardisierte Codierung von Zeichen wie der ASCII-Zeichensatz, Erweiterungen und Erweiterungstechniken, nicht-lateinische Alphabete, andere Sonderzeichensätze liegen als ISO Normen vor.
- Formeln (entweder mit erweiterten Zeichensätzen zu erfassen, mit speziellen Formelnotationen (siehe Algorithmen) oder als Graphik)
- Tabellen können mit Sprachmitteln beschrieben oder als Abbildung erfaßt werden.
- Graphiken können durch Sonderzeichenkombinationen erzeugt, sprachlich beschrieben (mit entsprechenden Computersprachen) oder mit Scannern erfaßt als "bitmap" vorliegen. Rastergraphik (oder Pixelgraphik) und Vektorgraphik sind entsprechende Unterscheidungen für die letztgenannten Typen. Auch Zeichen eines Alphabets können, wenn ihre *typographische* Bestimmung mit einbezogen wird, wie Graphik behandelt werden.
- Halbtonabbildungen wiederum sind wie scannererfaßte Graphiken zu betrachten.⁵⁴

⁵² vgl. "Elektronisches Publizieren im Rahmen der Textkommunikation" auf Seite 23

⁵³ ISO 646 ist praktisch mit dem ASCII-Zeichensatz identisch; Erweiterungen für verschiedene Sprachen und Symbolsätze arbeiten vor allem mit ISO 6937 (Coded Character Sets for Text Communications) und ISO/DIS 2022 (ISO 7-Bit and 8-Bit Coded Character Sets - Code Extension Techniques).

⁵⁴ Software, Programme (Algorithmen), Formel-, Tabellennotationen und Markup stellen in gewisser Weise eigene "Textelemente" dar, deren Spezifika hier nicht weiter verfolgt werden. Solange sie nicht unter dem Gesichtspunkt der Manuskripterstellung betrachtet werden, nicht unter dem Gesichtspunkt der Verarbeitung, verhalten sie sich wie normaler Text.

3.3 ENTWICKLUNGSSCHRITTE ZUR SGML

Das Elektronische Publizieren hat diese *integrierte, standardisierte* Verbindung offener Systeme mit differenzierten Anwendungsschichten und unterschiedlichen Präsentationsformen "idealtypisch" vor Augen, auch wenn selten explizit auf das ISO-Referenzmodell Bezug genommen wird.⁵⁵

Vom Publikationsprozeß aus gesehen sind die drei genannten Prozesse Manuskripterstellung, Typoskripterstellung und Ausgabeaufbereitung Stufen der Dokumenterstellung, für die je spezifisch Computerunterstützung erfolgen kann.

Im Zusammenhang der Standard Generalized Markup Language interessiert zunächst die erste Stufe und ihr "Zwischenprodukt", das elektronische Manuskript, das dann Ausgangsmaterial für die weiteren "Veredelungsstufen" ist. Wie bereits angeschnitten (vgl. "Die Beziehung Autor - Verlag - Satzbetrieb" auf Seite 29) schafft die zunehmende Zahl von Autoren, die ihre Manuskripte mit Textverarbeitungssystemen auf Mikrocomputern erfassen, eine neue Situation und für die Verleger die Chance, die Neuerfassung der Manuskripte (*u.U.*) zu vermeiden, die weitere Verarbeitung des Textes zu vereinfachen und die Verwertungsmöglichkeiten zu vervielfältigen.

Wenn beim Autor anfallende maschinenlesbare Daten als Manuskript an den Verlag gehen, stellt sich i.d.R. heraus, daß die Autorensysteme (bzgl. Speichermedium, Diskettenformat und Steuer codes der Textverarbeitungsprogramme) mit den Satzrechnern erst einmal nicht kompatibel sind.⁵⁶

Das Konzept der Text- oder, um Mißverständnissen vorzubeugen, der Dokumentverarbeitung nach Maßgabe eines internationalen Standards, der *wesentlich der Austausch- und Weiterverarbeitbarkeit* von Manuskripten zwischen und in Offenen Systemen dienen soll, basiert auf einer Reihe von Prinzipien, die sich aus der Kritik bestehender

⁵⁵ In der Literatur habe ich nur bei Petersen diesen Bezug direkt gefunden, obwohl gerade der Bezug auf das Modell helfen könnte, die anstehenden Probleme zu systematisieren. Üblich ist es, entweder über Dokumentaustausch oder die Kommunikation in Offenen Systemen zu sprechen. Für das integrierte Elektronische Publizieren jedoch müssen Geräteschnittstellen und Datenformate genormt, sprich kompatibel, sein.

⁵⁶ Das Konvertierproblem und das Autor-Verlags-Verhältnis wurde in "Fremddatenübernahme" auf Seite 33 erörtert.

Textverarbeitungssysteme entwickeln lassen. Die Entwicklung bzw. die Mängelbehebung werde ich kurz mit Hinweisen auf konkrete Systeme deutlich machen.

1. Bei Textverarbeitungssystemen für Mikrocomputer, vom Typ WordStar, um ein verbreitetes System zu nennen, muß ein Autor - zusätzlich zum Inhalt - spezielle Steuerbefehle (Control- und/oder Escape-Sequenzen) in den Text einfügen. Darüber wird dem Rechner mitgeteilt, wie bestimmte Textteile für die Ausgabe auf angeschlossenen Druckern zu verarbeiten sind. Bei dieser Form, die Zusatzinformationen in den Text einzustreuen, spricht man von *darstellungssorientierter* Auszeichnung (presentation Markup). Unter einem anderen Gesichtspunkt kann auch von *verarbeitungsorientierter* Auszeichnung (procedural Markup) /vgl. Goldfarb 1985, 133/ gesprochen werden. Die Markierungen, die in den WordStar-Text eingefügt sind, beziehen sich nämlich einerseits auf die Gestaltung des Textes und sind andererseits Steuercodes für angeschlossene Ausgabegeräte.

Was von einem austauschfähigen elektronischen Manuskript gefordert wird, daß es *keine* Informationen über Gestaltung der Ausgabe und Ausgabegerät enthalten soll, wird hier verletzt. Programme und Daten sind nicht getrennt. Der Text ist eine Mischung aus Programmanweisungen und Inhalt. Die Markierungen im Text sind direkte Befehle. In anderem Zusammenhang spräche man davon, daß Daten und Programme nicht getrennt sind. Bei dieser Art von Markup ist auch nicht zu unterscheiden, ob Textteile logisch oder layoutbezogen ausgezeichnet werden. Im Vordergrund jedenfalls steht noch das Präsentations-Markup, auch wenn damit zusammenfallend logische Elemente ausgezeichnet werden. Z.B. kann am Ende jedes Paragraphen eine Leerzeile durch einen Befehl angefordert werden: eine darstellungsbezogene Auszeichnung, die einem logischen Element gilt.

Die Nachteile liegen auf der Hand: die Befehle des Textverarbeitungssystems sind nicht die des Satzrechners und nicht die eines anderen Textsystems; selbst wenn eine Konvertiererroutine existierte,⁵⁷ ist höchst fraglich, ob die Gestaltung, die von einem Autor z.B. für seinen Matrixdrucker gewählt wurde, den Vorstellungen und Möglichkeiten eines Verlags entspricht.

2. In einer Weiterentwicklung von Textverarbeitungssystemen hat man die direkte Gerätesteuerung aus dem Text herausgenommen und

⁵⁷ WordStar stellt noch einen günstigen Fall dar, weil wegen seiner weiten Verbreitung für viele Zwecke Konvertierprogramme geschrieben worden sind.

in spezielle Programmteile verlegt. Mit anderen Worten, man hat Formatierprogramme geschrieben. Im Text werden nur noch Marken gesetzt, die speziell zu verarbeitende Textteile kenntlich machen. Die Befehle oder Befehlssequenzen werden nur noch vom Text aus aufgerufen, liegen selbst aber auf einer anderen Ebene. Damit einher geht zunächst - innerhalb eines geschlossenen Systems -, daß ein Text z.B. auch für den Bildschirm formatiert werden kann, bevor er für den Druck bestimmt wird und etwa wahlweise ein Matrixdrucker oder Typenraddrucker angeschlossen werden kann (Omniword könnte hier als Beispiel herangezogen werden).

3. Wenn die Formatierer erlauben, verschiedene Befehlssequenzen wahlweise mit den Aufrufen zu verknüpfen, d.h. wenn die Formatierer über eigene Makrosprachen verfügen, ist damit weitere Flexibilität gewonnen. Zu den verschiedenen Markierungen können je nach gewünschter Präsentationsform und zur Verfügung stehendem Ausgabegerät spezielle "Makros" aufgerufen werden. So ist ein Text für viele Ausgabegeräte und Darstellungsformen zu benutzen, ohne selbst verändert werden zu müssen. Mit der erreichten Flexibilität ist aber noch keine Portabilität gewonnen, da Textbearbeitungssystem und Formatierer noch ein geschlossenes System bilden. Nachteilig ist die Vermischung von gestaltbezogener und logischer Auszeichnung, wenn die Texte für den Druck *und* für die Übernahme in ein Retrievalsystem aufbereitet werden sollen.
Sollen z.B. alle Überschriften zur Hervorhebung unterstrichen werden, wird man die entsprechenden Marken für "Unterstreichung ein" und "Unterstreichung aus" setzen. Aber auch andere Teile im Text wird man für eine Unterstreichung markieren wollen. Damit werden die Markierungen unbrauchbar, um z.B. Überschriften in ein Feld Überschriften einer Datenbank einzulesen, weil die Marken ihre Eindeutigkeit verloren haben. *Logische* Auszeichnung dagegen ließe zu, daß Überschriften und andere Texte verschieden markiert, für den Druck aber gleich gestaltet werden.
4. Im nächsten Schritt, der vom vorhergehenden die Trennung von Textauszeichnung und Verarbeitung übernimmt, wird die "Philosophie" der Textauszeichnung grundlegend verändert. Statt der gestaltbezogenen oder formatierbezogenen Markierungen werden *nur noch* die logischen Elemente eines Dokuments markiert; logisches und gestaltungsorientiertes Markup werden entmischt. Die Gestaltung der Elemente wird den Formatierprogrammen überlassen. "Der Benutzer wird somit von der Problematik des Formatierens befreit und kann sich auf die logische Strukturierung

seines Textes konzentrieren /Scheller, Smith 1986, 1/.⁵⁸ Statt seinen Text selbst zu gestalten, wird seine Aufgabe nun, die Dokumentelemente bzw. die Dokumentstruktur zu kennzeichnen. Man spricht von logischer Auszeichnung in Hinblick auf die logische (hierarchisch, sequentielle) Struktur eines Dokuments und die Kennzeichnung der Knotenpunkte in dem Strukturbaum. Von deskriptivem Markup kann im selben Sinn mit einer kleinen Akzentverschiebung gesprochen werden, weil die Elemente der logische Struktur eines Dokuments mit mnemonischen Codes beschrieben (!) werden (z.B.: p = Paragraph; fn = Fußnote etc). Dies Konzept ist im angelsächsischen Raum zuerst als "Generic Coding" bekannt geworden, was sich nur schlecht mit "generischem Codieren" übersetzen läßt - aber doch noch wörtlichen Übersetzungen vorzuziehen ist. In der Einführung zur ISO/DIS 8879 wird dieser Schritt gegenüber der früheren Stufe hervorgehoben":

"While the macro calls could be placed anywhere in a document, users began to recognize that most were placed at the start or end of document elements. It was ,therefore, natural to choose names for such macros that were 'generic identifiers' of the element types, rather than names that suggested particular processing (e.g. 'heading' rather than 'format-17'), and so the practice of 'generic coding' (or 'generalized tagging') began."
/ISO/DIS 8879 1985, 1/

Vorreiter dieser Entwicklung war die Graphic Communications Association in den USA, die auch späterhin die Entwicklung der SGML unterstützte. Die GCA entwickelte ein generisches Codierschema, GenCode (eingetragenes Warenzeichen), das "many publishers in the USA - including the influential Department of Defense of the US Government" /Cave 1985, 6/ übernahmen. GenCode wurde auch der ISO zur Normung vorgelegt und ist in die Arbeit an der SGML eingeflossen.⁵⁹

Den Stand, der mit dem "generischen" Codieren erreicht ist, faßt Goldfarb knapp zusammen:

⁵⁸ Zur Diskussion steht freilich, ob diese "Befreiung" nicht auch zu Mehrbelastungen führen kann und zum Verlust von "Gestaltungspotential". Weiter unten werden die Akzeptanzbedingungen der Dokumentverarbeitung wieder aufgegriffen.

⁵⁹ Mehr zu GenCode bei /Bigman 1982/

In generic coding schemes, recognition, mapping, and processing can be accomplished all at once by the simple device of using GIs (Generic Identifier - K.B.) as control procedure names. Different formats can then be obtained from the same markup by invoking a different set of homonymous procedures. This approach is effective enough that one notable implementation, the SCRIBE system, is able to prohibit procedural markup completely. /Goldfarb 1985, 135/

5. Aufbauend auf dem Prinzip des deskriptiven Markup wurde Ende der siebziger Jahre begonnen, Generalized Markup Languages, schlecht übersetzt als allgemeine Auszeichnungssprachen, zu entwickeln. Das "generalized" ist auf ihre breite (verallgemeinerte) Verwendbarkeit für verschiedenste "Medien", Anwendungen und Ausgabegeräte zu beziehen. "Neutrale Auszeichnungssprache" machte im Deutschen noch klarer, worum es geht. Der Terminus Sprache deutet es schon an, daß hier mehr als ein Codierschema versprochen wird. Als unerledigte Probleme des "Generic Coding" konnten zum Beispiel ein großer Schreibaufwand besonders bei geschachtelten Elementstrukturen (z.B. Listen in Listen etc.) oder die Unmöglichkeit, Verweise auf Textteile (Kapitelüberschriften, Literaturverweise u.a.) zu erzeugen, angesehen werden.

Die Idee einer Sprache für die Dokumentverarbeitung will nicht nur diese Mängel beseitigen, sondern die Dokumentverarbeitung quasi auf den Stand der Verarbeitung "wohldefinierter Objekte" heben. In dem Ausdruck vom "rigorosen" Markup wird diese Zielsetzung zusammengefaßt, von C.F. Goldfarb wiederum, einem der "Gründerväter" der Generalized Markup Language⁶⁰ und selbst an der Entwicklung der DCFGML (Document Composition Facility Generalized Markup Language)/vgl. Goldfarb 1984/, einem IBM-Produkt, beteiligt:

"Markup should be rigorous so that the techniques available for processing rigorously-defined objects like programs and data bases can be used for processing documents as well"/Goldfarb 1985, S. 134/.

Dieser Schritt, der das "Generic Coding" oder deskriptive Markup als wesentlichen Zug einer allgemeinen Auszeichnungssprache bereits voraussetzt, konkretisiert sich in Dokumenttypdefini-

⁶⁰ Als frühen Artikel zu dem Ansatz siehe /Goldfarb, Mosher, Petersen 1970/

tionen (document type definition), die in gewissem Sinn den Deklarationsteilen bei Programmiersprachen ähneln: "Type definitions include a specification (like a formal grammar) of which elements and attributes can occur in a document and in which order"/ISO/DIS 8879, 1985, S.2/.⁶¹ Es wird also definiert, welche Elemente in bestimmten Dokumenten vorkommen können, welche Attribute ihnen zukommen und welche Werte sie annehmen können (diese Qualifizierung geht über die Möglichkeiten des 'Generic Coding' hinaus) und in welcher Reihenfolge (z.B. Berücksichtigung von Schachtelungen) sie vorkommen müssen und welche Datentypen für die Inhalte zulässig sind (i.d.R. beliebige Zeichenketten). Ein Dokument kann also, wenn eine Implementation einer Markup-Sprache vorliegt, ähnlich wie mit einem Compiler prozessiert werden, es können Syntaxfehler festgestellt werden, und ggf. kann die Auszeichnungstätigkeit aus dem "Wissen" um die Dokumenttypdefinition ergänzt werden.

Ein weiteres Charakteristikum der Dokumentverarbeitung mit einer Markup-Sprache, das dem Aufruf von Prozeduren oder Unterprogrammen in der gewohnten Programmierung ähnelt, sind spezifische "Referenzen". Mit diesen Verweisen können aus dem Dokument heraus Teile des Dokuments, die logisch und physikalisch als selbständige Dateien gespeichert sind, beim Prozessieren aufgerufen und in das aufrufende Dokument eingefügt und abgearbeitet werden.

Ein prominentes Beispiel für die Implementation einer "Generalized Markup Language" bietet die "Document Composition Facility" (DCF) von IBM. "The IBM Document Composition Facility is a text processing program; its main component is the text formatter, called SCRIPT/VS. One of the major features of SCRIPT/VS is the ability to develop and use the Generalized Markup Language (GML)"/Perry 1985, iii/. Die Ausgangsbasis ist hier ein Formatierer. Eine Möglichkeit, Dokumente herzustellen, ist, SCRIPT/VS-Formatierbefehle in den Text einzustreuen. Soweit hat die DCF nichts mit GML zu tun. Die andere Variante der DCF arbeitet mit der GML.⁶² Interessant ist die GML von IBM, weil sie viele Charakteristika enthält, die sich in der SGML wiederfinden. Die GML hat offensichtlich die Normungsaktivitäten der ISO in diesem Bereich stark beeinflusst. In einem Artikel zu verschiedenen Standardisierungsbemühungen zum Dokumentaus-

⁶¹ In der Beschreibung der SGML wird der Aufbau von Dokumenttypdefinitionen eingehend erklärt.

⁶² Zur DCF siehe auch /Document Composition Facility 1985/.

tausch bezeichnen Schindler, Flasche, Herrtwich /1985, 226/ die SGML als "important IBM product". Nun ist SGML nicht GML, der Einfluß der GML auf die SGML aber nicht zu leugnen.⁶³

"... it (SGML -KB) is based on the IBM product Document Composition Facility (DCF) Generalized Markup Language (GML). GML was developed by Charles F. Goldfarb in conjunction with two colleagues at the IBM Research Division, San Jose, California. The concepts on which it is based were taken into the American National Standards Institute (ANSI), and from thence to the International Organisation for Standardization (ISO), so an international standard could be developed." /Smith 1986 c, 193/

Auf Basis eines komplexen Formatierprogramms wird die GML "nachträglich" als Markup-Benutzeroberfläche entwickelt. Sie soll einfacheres und schnelleres Markup dadurch ermöglichen, daß die Elementnamen leicht zu erinnern sind und weniger Zeichen benötigen als die "Punktbefehle" des Formatierers.⁶⁴

SCRIPT/VS ist einer der flexiblen Formatierer, die bereits erwähnt wurden. Die Software enthält eine eigene Makrosprache, mit der Abfolgen von Formatierkommandos bestimmten Aufrufen oder Namen zugeordnet werden können. Das ist die erste Eigenschaft, die sich die GML zunutze macht. Nicht Formatierbefehle stehen im Text, sondern deskriptives Markup, das bei der Verarbeitung Makros aufruft. Es ist auch möglich, die Auszeichnung eines logischen Elements mit verschiedenen Formatierbefehlsfolgen (Makros), in der IBM-Terminologie "application processing functions" APFs, zu verbinden. Makros, die zu einer be-

⁶³ Andererseits, darauf machen die Autoren aufmerksam, konkurriert die DIA (Document Interchange Architecture) von IBM mit der SGML. Es ist nicht leicht zu sagen, ob IBM eine Harmonisierung betreiben, beide Produkte parallel laufen lassen, oder sich ab einem bestimmten Zeitpunkt für eine "Produktlinie" entscheiden wird.

⁶⁴ Das "Aufgesetzte" merkt man der DCFGML noch an. So werden Fehlermeldungen, die sich auf Verletzungen der Syntax der zugrundeliegenden Dokumenttypdefinition beziehen, zum Teil noch auf Ebene des Formatierers abgehandelt. Das bedeutet, daß ein GML-Nutzer seine Fehlermeldungen nicht in Termini der GML, sondern von SCRIPT/VS gemeldet bekommt. Somit kommt er nicht umhin, sich mit den Formatierbefehlen zu beschäftigen, was gerade durch GML vermieden werden sollte.

stimmten Anwendung gehören, können zu Profilen zusammengefaßt, in "Bibliotheken" abgelegt und wahlweise aufgerufen werden.⁶⁵

Auf der Ebene der Dokumentbeschreibung, und damit kommt das "rigorose" Markup ins Spiel, wird ein "allgemeines Dokument", eine Dokumenttypbeschreibung "allgemeines Dokument", definiert. Es wird ein bestimmter Satz von logischen Elementen für das "allgemeine Dokument" definiert. Die Elementnamen, die "generic identifier", werden bei der Texterfassung dem Text hinzugefügt. In der Wiedergabe der Dokumentstruktur auf der nächsten Seite sind die spezifischen logischen Elemente des "allgemeinen Dokuments" und die "GIs" in spitze Klammern eingeschlossen erfaßt.

Die Überschriften als Gliederungselemente sind hier den logischen Elementen zugeschlagen. Zu den Grundelementen, die offensichtlich keinen festen Platz in der Dokumentstruktur einnehmen (müssen) und die die Elemente sind, in die Text "eingefüllt" wird (zwischen Anfangs- und Endemarke), zählen: verschiedene Listentypen (geordnet, ungeordnet, Definitionslisten), Fußnoten, Abbildungen, Beispiele, lange Zitate, Anmerkungen, Glossar, Absätze, Zitate, Querverweise, Hervorhebungen, Worte, einzelne Zeichen.

Es ist möglich, zu dem Dokumenttyp "allgemeines Dokument" andere spezifische Dokumenttypdefinitionen zu entwickeln.

Die Entwicklung, die nachgezeichnet wurde und zur SGML führt kann in wenigen "Prinzipien" zusammengefaßt werden.

- Ein Dokument besteht aus Inhalt und Strukturen
- Logische Struktur und Layout-Struktur sind zu unterscheiden.
- Das Manuskript besteht aus Inhalt und logischer Struktur
- Die logische Struktur ist im Manuskript explizit zu beschreiben

⁶⁵ SCRIPT/VS kann außerdem das Markup in die Steuerzeichen anderer Textverarbeitungsprogramme konvertieren, z.B. für Satzsysteme. Es können Informationen aus dem Dokument für den Aufbau von Datenbanken extrahiert werden, oder bestimmte Inhalte können aus Datenbeständen wiedergewonnen werden.

```

allgemeines Dokument <gdoc>
  Vorspann <frontm>
    Titelseite <titlep>
      Dokumenttitel <title>
      Dokumentnummer <docnum>
      Datum <date>
      Name des Autors <author>
      Adresse des Autors <address>
        Adreßzeile <aline>
    Zusammenfassung <abstract>
      Grundelemente
    Vorwort <preface>
      Grundelemente
        Überschriften <h2> bis <h6>
      Grundelemente
    Inhaltsverzeichnis <toc>
    Abbildungsverzeichnis <figlist>
  Hauptteil <body>
    Teilüberschrift <h0>
      Grundelemente
    Kapitelüberschrift <h1>
      Grundelemente
        Überschriften <h2> bis <h6>
      Grundelemente
  Anhang <appendix>
    Überschrift <h1>
      Grundelemente
        Überschriften <h2> bis <h6>
      Grundelemente
  Schlußteil <backm>
    Überschrift <h1>
      Grundelemente
        Überschriften <h2> bis <h6>
      Grundelemente
  Index <index>

```

- Die Beschreibung der Struktur oder m.a.W. die logische Auszeichnung wird durch Markierungen im Manuskript (Markup) vorgenommen
- Die Verarbeitungsebene ist von der inhaltlichen & logischen Ebene (Manuskript) zu trennen: Trennung von Daten und Programmen
- Die Befehle zur Verarbeitung eines Dokuments werden an das deskriptive Markup, die logische Struktur geknüpft

- Layout-Struktur (i.w.S. Ausgabeformat) entsteht durch spezifische Verarbeitung der logischen Auszeichnungen
- Dokumente mit gleicher logischer Struktur bilden Dokumentklassen
- Dokumenttypdefinitionen legen gültiges Markup für Dokumente einer Klasse fest
- Dokumenttypdefinitionen können die Eingabe korrekten Markups unterstützen und kontrollieren
- Die Regeln für die Konstruktion von Auszeichnungsschemata und Dokumenttypdefinitionen können in einer Metasprache beschrieben und standardisiert werden: Standard Generalized Markup Language

4. ELEKTRONISCHES PUBLIZIEREN MIT DER SGML

In diesem Kapitel wird die Standard Generalized Markup Language in einiger Ausführlichkeit beschrieben, weil sie möglicherweise (!) einen Dreh- und Angelpunkt für die weitere Entwicklung "elektronischer Manuskripte" und Elektronischen Publizierens insgesamt darstellen kann. Bereits jetzt - vor der Verabschiedung des "International Standard" - bildet die SGML einen beachtlichen Fokus für Projekte zum Elektronischen Publizieren.

4.1 DIE ISO/DIS 8879 (SGML)

Nachdem im letzten Abschnitt des vorigen Kapitels das Konzept einer Standard Generalized Markup Language vor allem unter dem Gesichtspunkt der Mängelbehebung entwickelt wurde, werden im ersten Abschnitt dieses Kapitels der Anspruch der Norm und die Vorteile, die von ihr für den Publikationsprozeß erhofft werden, zusammengestellt. Im nächsten Abschnitt stelle ich Entwicklung und aktuellen Stand der Normung dar. Danach erläutere ich die Grundzüge der SGML soweit, wie es für das Verständnis konkreter Anwendungen nötig ist.

4.1.1 Versprechungen der SGML

Die SGML bietet sich als Instrument für das Elektronische Publizieren in Hinblick auf die Portabilität und Weiterverarbeitbarkeit von Texten und speziell für die Weiterverarbeitung für Mehrfachnutzungen an.⁶⁶

⁶⁶ "The Standard Generalized Markup Language can be used for documents that are processed by any text processing or word processing system. It is particularly applicable to:

1. Documents that are interchanged among systems with different processing languages.
2. Documents that are processed by more than one application, even when the applications use the same text processing language." /ISO/DIS 8879, 1985, 5/

"SGML can be used for publishing in its broadest definition, ranging from single medium conventional publishing to multimedia data base publishing. SGML can also be used in office document processing when the benefits of human reliability and interchange with publishing systems are required."
/ISO/DIS 8879, 1985, 1/

Die SGML wird als ein Instrument angesehen, das im Bereich des kommerziellen Publizierens und im wachsenden Bereich des Inhouse-Publishing (vor allem Organisationen, die Technische Dokumentationen oder amtliche Dokumentationen auf dem Laufenden zu halten haben) von Vorteil ist.

Für den kommerziellen Verleger eröffnet die SGML vier positive Perspektiven.

1. Der erste Vorteil entsteht, wenn der Verlag ein elektronisches Manuskript übernehmen kann, das bereits ein SGML-Dokument ist. Dann entfallen natürlich die Kosten für die Neuerfassung oder aufwendige Konvertierungen. Im Extremfall wäre das Manuskript so perfekt mit SGML-Markup versehen, daß es automatisch gesetzt werden könnte. Es ist natürlich eine offene Frage, wer das Markup in den Text bringt: der Autor, der Verlag (bzw. die Redaktion) oder beide zu gewissen Anteilen? Die Last, die damit verbunden ist, ist allerdings auch abhängig von der für das Markup zur Verfügung stehenden Software. Vorteile ergeben sich, wenn man die Frage des Lastenausgleichs einen Moment beiseite läßt, für beide Seiten. Für den Autor und den Verlag ist ein Interesse an schneller Veröffentlichung gegeben. Für beide Seiten ist zudem die gegenwärtige Situation des "ragbag approach", wie Joan Smith das formuliert, unbefriedigend. Der "ragbag approach", d.h. das Zusammenstellen von Typoskripten unterschiedlichster Qualitäten, Schriften und Layouts zu Sammelbänden führt zwar zu schneller Lieferung, spart auch die Satzkosten, erhöht aber die Papierkosten (weil gesetzter Text platzsparend *und* gut lesbar ist) und senkt die Lesbarkeit, von ästhetischen Gesichtspunkten ganz abgesehen.
2. Beim Verlag entsteht durch die Übernahme maschinenlesbarer Manuskripte eine Datenbasis, die genutzt werden kann, und zwar auf eine besonders einfache und vielfältige Weise, wenn die Dokumente der Datenbasis SGML-Dokumente sind. Zunächst gestattet eine Datenbasis ständiges "Updating" und somit die unverzügliche Herausgabe aktualisierter Auflagen.

Die SGML-Datenbasis erlaubt weiter verschiedene Ausgabestile, unterschiedlich etwa für eine Korrekturvorgabe oder eine "Lu-

xusausgabe" oder verschiedene Ausgabevarianten, etwa für eine englische und eine amerikanische Version.

Aus der Datenbasis - immer noch im Bereich der Printmedien - können neue Zusammenstellungen von Texten zu Publikationen entstehen. Aus den letzten Jahrgängen von zehn Zeitschriften zur Informationswissenschaft könnte z.B. ein Sammelband zum Elektronischen Publizieren gewonnen werden, der das Erscheinungsbild der Artikel gänzlich änderte. Eine im Prinzip ähnliche Anwendung hebt J. Smith in bezug auf Nachschlagewerke hervor. Sachbezogene oder historische Selektionskriterien könnten auch hier neue Publikationen hervorbringen /vgl. Smith 1986 a, 3/. Eine weitere Anwendung wäre, Abstracts aus Zeitschriftenartikeln und bibliographische Angaben zu einem Referatedienst zusammenzustellen. Für alle diese Publikationsformen, die aus einer gemeinsamen Datenbasis gewonnen werden, gilt, daß sie je spezifisch gestaltet werden können.

3. Da die SGML-Dokumente ohne Probleme (immer prinzipiell in der EDV) in ein Retrievalsystem eingespielt werden können, kann die verlagsinterne Datenbank auch den Weg an die Öffentlichkeit antreten. Mit der Suche in der Wissensbasis nach Informationen kann entweder eine traditionelle Bestellung einer Publikation veranlaßt werden oder ein "printing on demand". Als dritte Möglichkeit, kann der Volltext noch über Telekommunikationswege an den Nachfrager gelangen. Hier liegt ein Entscheidungsbereich für den Verleger, ob er Publikationen parallel anbieten soll oder nur noch in elektronischer Form (electronic journal), gedruckte Auflagen nicht mehr vorhalten, sondern nur noch auf Bestellung drucken und liefern soll. Die Entscheidungen stehen im Kontext der nachlassenden Nachfrage der Bibliotheken und der Einbußen durch Kopierzentren.
4. Als weitere Einnahmequelle und Aktivität kann ein Verleger seine Daten für Mikrofilm, Videotex und CD-ROM aufbereiten und publizieren.

Für den "Inhouse-Publisher", der vor allem sich verändernde Dokumentationen aktuell halten muß (Beispiel Softwaredokumentation), ist eine maschinenlesbare Datenbasis vorteilhaft. Die SGML erlaubt zusätzlich, die Dokumente gleich zu gestalten, unabhängig davon, wer sie erfaßt. Die SGML erlaubt darüber hinaus noch Erleichterungen bei der Erfassung, z.B. Einsparung von Anschlägen. Das SGML-Format hält die Inhouse-Datenbank für den Kontakt mit kommerziellen Verlegern oder Setzern offen.

4.1.2 Zu Geschichte und Stand

Die SGML, wie sie im Draft International Standard ISO/DIS 8879 "Information processing - Text and office systems - Standard Generalized Markup Language (SGML)" festgelegt ist - mir liegt der Normentwurf vom 10.10.1985 vor, zu dem bis zum 10.4.1986 noch Änderungsvorschläge eingebracht werden konnten - hat schon Vorgeschichte. Dem jetzigen Entwurf sind neun Arbeitsentwürfe und ein Entwurfsvorschlag vorausgegangen. Die Entwicklung zur SGML begann bei der ISO 1979 im "Computer Languages for the Processing of Text development project" im TC 97/SC 5, wurde aber später an das SC 18 abgegeben.⁶⁷ Ursprünglich war ein mehrteiliger Standard vorgesehen worden, der z.B. noch Verbindungen zum GKS (Graphisches Kernsystem) oder Anwendungen in WYSIWYG (What You See Is What You Get)-Systemen behandeln sollte. Der Teil sechs, der die SGML enthält, wurde dann später aus dem Gesamtkomplex herausgenommen, und im April 1985 wurde beschlossen, SGML als unabhängige Norm weiterzuentwickeln /vgl. ISO/DIS 8879 1985, 163/. Der Entwurf, der mir vorliegt, enthält neben dem eigentlichen Normtext noch eine Reihe von Anhängen, die zum Teil einführenden und zum Teil kommentierenden Charakter haben, selbst aber nicht zur Norm gehören.⁶⁸

Der Gang der Entwicklung hat, wie schon gesagt, bei der DCF von IBM begonnen, ist dann zum amerikanischen Norminstitut und von da weiter in die ISO getragen worden. Im Lauf der Normungsarbeiten ist das

⁶⁷ TC 97 = Technical committee
97 (Information processing systems)
SC 5 = Subcommittee
5 (Computer Languages)
SC 18 = Subcommittee
18 (Text and office systems)
WG 8 = Working Group
8 (text processing and markup languages)

⁶⁸ Der Annex A "Introduction to Generalized Markup" ist mit dem bereits zitierten Artikel von C. F. Goldfarb /Goldfarb 1985/ identisch. Annex B "Basic Concepts", Annex C "Additional Concepts", Annex D "Public Markup Constructs" erläutern die Norm. In Annex E werden Anwendungsbeispiele gegeben, in Annex F "Implementation Considerations" angestellt. Anhang G und H befassen sich kurz (jeweils weniger als zwei Seiten) mit "Theoretical Basis for the SGML Content Model" und "Nonconforming Variations". /Vgl. ISO/DIS 8879/

Ausgangsmaterial so verändert worden, daß die DCFGML keine mit der Norm konforme Implementation der SGML darstellte.

Der Status als Draft International Proposal bedeutet, daß die SGML noch ein Stück Weges bis zum Internationalen Standard vor sich hat. Im April 1986 hat sich die zuständige WG 8 des ISO/TC97/SC 18 getroffen, um die zur ISO/DIS 8879 eingegangenen Änderungsvorschläge und Voten auszuwerten.⁶⁹ Fünfzehn Länder stimmten ab. Tschechoslowakei, Frankreich, Ungarn, Italien, Niederlande, Norwegen und UdSSR stimmten für die Norm; Österreich, Kanada, Bundesrepublik Deutschland und die USA stimmten zu, aber mit Kommentar. Irland, Polen und die Schweiz enthielten sich, und nur Großbritannien stimmte mit Kommentar dagegen. Damit ist der Wahlgang für den DIS erfolgreich gewesen. Wegen der zahlreichen Kommentare ist mit noch erheblichen Änderungen für die Endfassung zu rechnen.⁷⁰ Die Änderungsvorschläge und die Entscheidungen über ihre weitere Berücksichtigung sind in einem Bericht des Normherausgebers (editors) (Charles Goldfarb) festgehalten. Der nächste Schritt ist nun, daß der Herausgeber auf Grundlage des Berichts den Text überarbeiten wird und ihn bis zum 15. September dem Sekretariat des SC 18 vorlegen wird, das ihn dann an das zentrale Sekretariat weiterleitet und den Wahlgang eröffnet. Die Mitglieder des ISO-Rates bekommen eine Kopie des Textes. Wenn innerhalb einer Frist von sechs Wochen kein negatives Votum eingeht, wird "stilles Einvernehmen" angenommen, und die Norm gilt. Ob die Norm dann akzeptiert wird, steht auf einem anderen Blatt, da ihr ja keine Rechtskraft zukommt.⁷¹ Für den Erfolg einer Norm ist sicherlich auch wichtig, welche Maßnahmen zu ihrer Unterstützung ergriffen werden und in welchem Normenkontext sie eingelassen ist.

⁶⁹ Die Informationen darüber stammen aus /Smith 1986 c/

⁷⁰ Auf der Markup '86 gab Charles Goldfarb einen Überblick über den derzeitigen Stand der SGML. Vor allem im Bereich der Anbindung der logischen Dokumentstruktur an Layoutstrukturen wick die Darstellung vom Text des Draft International Standard ab. In sog. "link type definitions" werden Entsprechungen zwischen logischer Struktur und verarbeitungsorientierten Strukturen hergestellt. Es ist möglich, daß dadurch eine Harmonisierung mit ODA hergestellt werden soll.

⁷¹ Es wäre einmal interessant zu sehen, welche großen "flops" bezüglich Akzeptanz in der Geschichte der ISO vorgekommen sind, um daraus zu lernen, was über Akzeptanz und Ignoranz von Normen entscheiden kann.

Neben der Entwicklung von Normen kann die ISO noch "flankierend" Technische Reports anfertigen. Als Technische Reports, zur Unterstützung der SGML, sind ein "Reference Model for Text Description and Processing Languages" und ein Bericht zur Terminologie "Computer-Assisted Publishing - Vocabulary" als "Draft Technical Report" auf den Weg geschickt worden. Weitere praktische Hilfe soll in einem Bericht "Techniques for Using SGML" versammelt werden. In diesem Bericht soll u.a. die Integration von Graphikelementen in SGML beschrieben werden und ein "starter document type" als Starthilfe für SGML-Anwender bereitgestellt werden (analog zu dem "allgemeinen Dokument" in dem IBM-Produkt GML) /vgl. Smith, Stutely 1986/. Diesen Bericht betreut J. Smith.

Im SGML-Norm-Kontext ist zuerst SDIF (SGML Document Interchange Format) relevant. Diese Norm befindet sich noch im Zustand des "Draft Proposal" und wird ebenfalls von Charles Goldfarb betreut. Ihr Zweck ist, SGML-Dokumente als Datenstrom für den Austausch in Offenen Systemen zu organisieren. Die Definition der SDIF erfolgt in der "Abstract Syntax Notation One", "to make it easier to lock into the standards for Open Systems Interconnection (OSI)" /Smith 1986 c, 2/.

Im Rahmen der WG 8 sind noch weitere Normen in Vorbereitung. Im Status eines "Working Draft" befinden sich "Requirements for SGML-Sensitive Text-Entry Systems", worunter Anforderungen an Editoren, die speziell auf die Erstellung von SGML-Dokumenten zugeschnitten sein sollen, zu verstehen sind. Im Jargon wird von "SGML-smart-editors" gesprochen. Daneben gibt es wieder einen Technischen Report, der die Arbeit mit SGML an anderen Erfassungsgeräten zum Inhalt haben soll ("Using SGML with Non-SGML Text Entry Systems"). Da SGML keine Vorschriften für die Konstruktion von Dokumenttypen macht, benötigt sie so etwas wie ein Informationsbüro für Dokumenttypen, die außerhalb von Inhouse-Anwendungen bekannt und angewendet werden sollen. Eine Norm über "Registration Procedures for Public Constructs" wird zu diesem Zweck entwickelt.⁷² Auch wenn die meisten Vorhaben noch nicht sehr weit fortgeschritten sind, ist doch ein recht umfassendes Geflecht von unterstützenden Maßnahmen beabsichtigt. Eine Unterstützung anderer Art bekäme die Norm von der

⁷² Zwei weitere Normvorhaben der WG 8 "Text Composition Semantics" und "Text Presentation Metafile" sollen nur am Rande genannt werden, da mir über ihren Inhalt und genauen Zusammenhang mit SGML nichts bekannt ist.

ISO, wenn diese SGML für ihre eigenen Publikationen verwenden würde, was im Gespräch ist /vgl. Smith 1986 e, 193/.⁷³

4.1.3 Die Bauform der SGML

Wenn im folgenden die ISO-Norm dargestellt wird, so halte ich mich zunächst eng an den Normtext und den Annex B "Basic Concepts" /ISO/DIS 8879, 1985, 68ff/, der nicht zum Standard gehört, ihn aber zu explizieren versucht. Der Normtext selbst greift auf Produktionsregeln für formale Syntaxen zurück, um die SGML möglichst präzise zu definieren /vgl. ebd., 7/ und ist deshalb weniger "zitiertgeeignet".

4.1.3.1 Kriterien einer standardisierten Markupsprache.

Als Standard greift die Standard Generalized Markup Language die Prinzipien des generischen Codierens und der Markup-Sprachen auf und standardisiert sie.⁷⁴ Ihr Anspruch ist gerade nicht, eine spezifische Implementation zu standardisieren, sondern als Metasprache /vgl. ISO/DIS 8879,3/ ein Regelwerk für Implementationen, die dem Standard genügen sollen, bereitzustellen.

Sie liefert drei Instrumentarien für die "Regelung" dreier Komplexe:

1. Sie enthält eine abstrakte Syntax, die die allgemeinen Regeln für das deskriptive Markup von Dokumenten enthält (also den Teil der Dokumentauszeichnung, der das generische Codieren regelt).
2. Sie enthält sprachliche Mittel, die für den Aufbau von Dokumenttypdefinitionen verwendet werden, d.h. festlegen, wie für bestimmte Dokumentklassen zulässiges Markup "deklariert" wird.
3. Sie legt eine konkrete Syntax, den englisch-sprachigen Raum vor Augen, die "reference concrete syntax", vor, und die Regeln nach

⁷³ Auch das DIN hat im Rahmen des DOCDEL-Programms mit SGML experimentiert.

⁷⁴ "The Standard Generalized Markup Language standardizes the application of the generic coding and generalized markup concepts." /ISO/DIS 8879, 1985, 2/

denen sich alternative konkrete Syntaxen für eigene Zwecke normkonform entwerfen lassen.

Der Standard darf aber nicht nur eine Bauanleitung für kompatible oder leicht konvertierbare Markup-Sprachen sein, sondern muß so angelegt sein, daß einschränkende Merkmale existierender Geräte, Speicherformate und Zeichensätze sich nicht negativ auf die Erstellung von SGML-Dokumenten in unterschiedlichen "Umgebungen" und ihren Austausch auswirken.⁷⁵

Neun Kriterien werden in dieser Hinsicht aufgestellt:

1. In Übereinstimmung mit dem Standard ausgezeichnete Dokumente müssen auf vielen Textverarbeitungssystemen verarbeitet werden können. Gedacht ist wohl besonders daran, daß "verarbeitungsneutral" ausgezeichnete Texte auch auf einfachen Textverarbeitungssystemen verarbeitbar sein sollen. SGML-Dokumente sollen auch von einfachen "Formatierern" verarbeitet werden.⁷⁶
2. Jedes Texteingabesystem soll nach der Norm auszeichnen können. Die Eingabe ist durch einfaches Anschlagen (jedenfalls bei der concrete reference syntax) von Tasten möglich und erfordert keine speziellen Tasten zur Eingabe von Steuercodes oder Sonderzeichen.
3. Es darf keine Zeichensatzabhängigkeit geben. Die SGML ist nicht an bestimmte Zeichensätze gebunden. Nicht darstellbare Zeichen werden durch einfache Zeichenfolgen codiert.
4. Es darf keine (kaum) Abhängigkeit von Anwendungen, Systemen oder Geräten geben. Da das Markup deskriptiv ist, also keine

⁷⁵ "... to use it in a multiplicity of environments." /ISO/DIS 8879, 1985, 3/

⁷⁶ Eine einfache Form der Verarbeitung wäre z.B. über eine Konkordanzliste zu erreichen, die bestimmte Auszeichnungen mit Steuerzeichen verbindet. Über die Funktion "Suchen und Ersetzen" in ein kleines Programm eingebunden, könnte die Konvertierung durchgeführt werden.

Gedacht ist aber wohl auch daran, daß die Systeme, zwischen denen Texte ausgetauscht werden, beide über Implementationen der SGML verfügen - nur unterschiedlicher Komplexität. Dann könnte über Anpassungen im Deklarationsteil (dazu unten mehr) relativ leicht erreicht werden, daß SGML-Dokumente in beiden Systemen bearbeitet werden können.

Verarbeitungsanweisungen enthält, ist das der Fall. Falls doch - was in der Realität wohl wahrscheinlich ist - spezifische Steuerzeichen im Text vorhanden sein müssen, z.B. der Zugriff auf eine Datei mit einer bestimmten physischen Adresse, müssen diese vor dem Austausch der Texte gesondert kenntlich gemacht werden.

5. Nationale Sprachen werden gleich behandelt, weil es keine Festlegung auf bestimmte Zeichensätze gibt und Namen und Bezeichnungen vom Anwender bestimmt werden.
6. Die Sprache muß sich an Schreibmaschinen- und Textverarbeitungskonventionen anpassen können. Ein Text, der Leerzeilen und Anführungszeichen enthält, sollte in besonderen Fällen bereits als SGML-Text interpretierbar sein.
7. Die Auszeichnungssprache muß von physischer Datenorganisation und Datenfluß unabhängig sein. "Für" die SGML besteht ein Dokument aus einer oder mehreren Entitäten.⁷⁷
8. Ein SGML-Dokument muß, entsprechend markiert, Daten enthalten können, die nicht durch die SGML definiert und auch nicht von SGML-Implementationen verarbeitet werden können. Ausgezeichneter Text muß andersherum auch in andere Datentypen, entsprechend markiert, eingefügt werden können.
9. Markup muß von Menschen und Maschinen verstanden werden.

"The Standard Generalized Markup Language is intended as a suitable interface for keyboarding and interchange without preprocessors" /ISO/DIS 8879 1985, 4/.

D.h. hier ist daran gedacht, daß die Texteingabe plus Markup so vereinfacht werden kann, daß die Auszeichnung nicht durch Software unterstützt werden muß. Andererseits erlaubt die Struktur der SGML aber auch, die Textauszeichnung durch "intelligente" Editoren oder mit Hilfe angepaßter Textverarbeitungssysteme zu unterstützen.

⁷⁷ "The markup language has a virtual storage model in which documents consist of one or more storage entities, each of which is a sequence of characters. All real file access is handled by the processing system..." /ISO/DIS 8879, 1985, 4/.

4.1.3.2 Dokumenttypdefinition und SGML-Parser

Bevor ich auf die Markup-Sprache im einzelnen komme, muß vorweggeschickt werden, daß der Wert einer Markup-Sprache, mit der man Dokumenttypen definieren und die logische Struktur eines Dokuments beschreiben kann, darin liegt, daß sie zu Implementationen führt. Die SGML liefert die Regeln, wie die Struktur relevanter Dokumente in Dokumenttypdefinitionen beschrieben werden kann. Die Dokumenttypdefinition ist die Grundlage für eine Software, die überprüfen kann, ob Dokumente der Dokumenttypdefinition entsprechen. Man spricht dann von SGML-Systemen oder SGML-Parsern /Vgl. ISO/DIS 8879 1985, 55/.

"Generally speaking, a parser is a program used to determine the underlying structure and content of some input object (file, document etc.). More formally (in an SGML context), a parser checks that the tokens appearing in the input document occur in patterns that are permitted (by the rules of SGML and the description given by the document architect in the document type definition) and makes explicit the hierarchical structure of the incoming token stream by identifying which parts should be grouped together" /Heath 1986, 5/.

Ein SGML-Parser ist mithin eine Software, die Dokumenttypdefinitionen, die der SGML entsprechen, einlesen und interpretieren kann, derart daß ein ausgezeichnetes Dokument daraufhin überprüft werden kann, ob es der Dokumenttypdefinition genügt. Wenn Fehler entdeckt werden, sollte das in entsprechenden Fehlermeldungen resultieren.⁷⁸ Anzumerken ist noch, daß für die Auszeichnung eines Dokuments selbst kein SGML-Parser nötig ist. Wenn ein SGML-Parser für eine bestimmte Anwendung zugeschnitten in ein Texterfassungssystem integriert wird, kann das die Auszeichnungstätigkeit erheblich erleichtern. Vor der Weiterverarbeitung von SGML-Dokumenten ist eine Syntaxkontrolle unumgänglich. Am Ende dieses Abschnitts komme ich noch einmal auf SGML-Implementationen und -Parser zurück.

⁷⁸ Auf die theoretischen Grundlagen des "checking for conformance" bei der Automatentheorie wird in /ISO/DIS 8879 1985, 155/ nur kurz hingewiesen.

4.1.3.3 Das Dokumentmodell der SGML

Die SGML baut auf einem hierarchischen Dokumentmodell auf. Das Modell ist selbst nicht Teil des Standards, wird im Annex B aber kurz dargelegt.⁷⁹

"... a document is a logical construct called a *document element*, the top node of a tree of elements that make up the document's *content*" /ISO/DIS 8879, 1985, 68/.

Ein Element "Magisterarbeit" oder kurz "Arbeit" könnte Kapitel, Überschriften, Unterkapitel etc. enthalten. Die Abbildung auf der folgenden Seite gibt grob die Baumstruktur des Elements (Dokuments) "Arbeit" wieder.

An den Knotenenden befände sich der eigentliche Textinhalt, die Daten. Diese zunächst nur implizite logische Struktur wird durch das Markup, das die einzelnen Elemente voneinander absetzt und bestimmt, explizit gemacht. Das Markup bedeutet das Hinzufügen von Informationen zu einem Text bzw. die Darstellung impliziter Informationen. Es lassen sich in der Terminologie der GML /vgl. Document Composition Facility 1985, vii und 7f/ zwei Elementarten unterscheiden. Elemente, in die Text eingelassen werden kann und die nicht prinzipiell an eine bestimmte Stelle in der Dokumentordnung gebunden sind, können als Grundelemente (basic document elements), solche, denen die Strukturierung des Dokuments obliegt, die aber selbst unmittelbar keinen Text enthalten, als Gliederungselemente (basic document structures) bezeichnet werden.

Das Markup eines *Elements* besteht aus einer Anfangsmarke (start-tag) und einer Endmarke (end-tag). Die Marken beschreiben die *Attribute* eines Elements. Jedes Element wird mit einem Attribut versehen, das ihm seinen Namen gibt und seinen Typus bestimmt. Dies Attribut wird "generic identifier", kurz GI, genannt - als "Elementkennung" übersetzbar.⁸⁰

Die Elementkennungen können selbst noch weitere Attribute erhalten.

⁷⁹ Vgl. /Scheller, Appelt 1985, 15/ "Man hat also in SGML dasselbe hierarchische Dokumentenmodell zur Verfügung wie in ODA, das Modell ist jedoch nicht Bestandteil des Standards."

⁸⁰ "generic identifier : A string identifying an element type, found in the start-tag and end-tag delimiting an element, and defined in an element declaration." /Smith, Stutely 1986, 3/

Alle Attribute zusammen beschreiben die Struktur der Elemente (des Dokuments). Sie beschreiben, welche Elemente in welcher Reihenfolge wie oft und mit welchen Ausprägungen vorkommen.

Alle Dokumente, die dieselben Elemente in einer mehr oder weniger festen Ordnung verwenden, bilden einen bestimmten Typ. Oder andersherum: es können Dokumenttypen definiert werden, die festlegen, welche Struktur ein Einzeldokument der Klasse annehmen darf.

Eine Dokumenttypdefinition regelt:

1. welche Elementkennungen in einem Dokument des bestimmten Typs erlaubt sind
2. welche Attribute die einzelnen Elemente annehmen können
3. welche Unterelemente eine Elementkennung enthält
4. welche Datentypen in den Elementen auftreten dürfen⁸¹

Die Dokumenttypdefinitionen gehören praktisch in den Deklarations-
teil eines Dokuments, werden aber nach Möglichkeit nicht laufend
verändert, sondern erst, wenn neue Anwendungen anstehen. In der
Regel wird die Festlegung der Dokumenttypdefinitionen mit der
Festlegung von Verarbeitungsprozeduren für eben diese Definitionen
einhergehen. Die Festlegung der Verarbeitung geschieht in der Sy-
stemsprache.

Die Dokumenttypdefinitionen gehören zum Dokument, bedürfen ihrer-
seits sprachlicher Mittel, um deklariert werden zu können (Markup
Deklarationen).

Ein dritter Typ von Markup wird nötig, um "ausgelagerte" Dokument-
teile in das Hauptdokument zu integrieren. Nach der SGML kann ein
Dokument aus verschiedenen Entitäten (entities) zusammengesetzt
sein, die in verschiedenen Dateien abgelegt sein können. Die Kon-
sequenz aus dieser "verteilten" Dokumentorganisation ist, daß vom
Hauptdokument aus die einzelnen Teile angesprochen werden müssen.
Diese Verweise im Hauptdokument (entity references) sind selbst
eine Zusatzinformation, Markup. Die Entitäten, d.h. die zugelas-
senen ausgelagerten Teile des Dokuments, müssen ihrerseits wieder
durch Markup Deklarationen definiert sein.

Ein Dokument, das von einer SGML Implementation verarbeitet werden
soll, muß - wie dargestellt - wenigstens drei Formen von Markup

⁸¹ Vgl. /ISO/DIS 8879, 1985, S.68/

enthalten: den "Deklarationsteil" (markup declaration), die Elementkennungen und die Verweise auf "ausgelagerte" Entitäten.⁸²

Es soll wieder der Dokumenttyp "Arbeit" herangezogen werden, um exemplarisch zu zeigen, wie der Deklarationsteil gebildet werden kann.

Der Deklarationsteil kennt Elementdeklarationen, die wiederum die Attributsdefinitionen des Elements enthalten und die Werte, die sie annehmen können. Für jedes Element im Dokument werden zwei Parameter definiert: die zugehörige Elementkennung und ein Modell ihres Inhalts (content model). So könnte die Deklaration für das Element "Arbeit", den obersten logischen Knoten in der Dokumentstruktur, so aussehen:

```
<! ELEMENT Arbeit (Vorspann, Corpus, Nachspann) >
```

Die Elementkennung (hier: Arbeit) wird durch ein "Modell" definiert.

"A model is an instance of a *group*, which is a collection of connected members, called *tokens* /ISO/DIS 8879, 1985, 74/

Eine Gruppe wird, wie man sieht, in Klammern eingeschlossen. Die Gruppe kann aus verschiedenen "token" bestehen, die verschiedene Verbindungen untereinander eingehen können und deren Vorkommenshäufigkeit festgelegt werden kann.

Verbindungsoperatoren sind das Komma und das Fragezeichen. Das Komma bedeutet, daß ein token auf den anderen folgen muß; ein Fragezeichen bedeutet, daß ein token optional ist, also einmal oder keinmal auftreten kann.

Wenn "Arbeit" nicht unbedingt einen Nachspann enthalten muß, sähe die Elementdeklaration so aus:

```
<!ELEMENT Arbeit (Vorspann, Corpus, Nachspann?) >
```

Vom obersten Knoten ausgehend, werden weitere Elementdeklarationen für alle weiteren Stufen und die darin vorkommenden Kennungen ausgewiesen. Im Beispiel beim Element "Vorwort" angekommen, das nur noch Abschnitte (Paragraphen) enthalten soll, ergäbe sich als Elementdeklaration:

⁸² Obwohl es gegen die "reine Lehre" verstößt, können auch Verarbeitungsinstruktionen als vierter Markup-Typus im Text vorkommen, dann natürlich als systemabhängiges Markup.

<!ELEMENT Vorwort (p+) >

Der Operator Plus (+) bedeutet, daß ein Element einmal oder mehrmals vorkommen kann, der Operator Asterisk (*) läßt ein optionales und wiederholbares Element zu.

Die Operatoren können sich auch auf eine Gruppe von Elementen beziehen. Angenommen in "Vorwort" wären auch Fußnoten zugelassen, sähe die Deklaration so aus:

<!ELEMENT Vorwort (p | fn)* >

Der vertikale Strich zeigt eine ausschließende "Oderung" an (entweder, oder). Durch den Asterisk wird diese ausschließende Oderung quasi in eine Schleife eingebaut. Paragraph und Fußnote sind damit als alternative Elemente optional und wiederholbar.

Das Element Titelblatt könnte so definiert sein, daß Titel, Autor und Floskel darin vorkommen müssen, aber in beliebiger Reihenfolge. Das leistet das Ampersand (&) als Operator.

<!ELEMENT Titelblatt (Titel & Autor & Floskel) >⁸³

Auf der Ebene des Textes, der Grundelemente angelangt, werden die zulässigen Daten deklariert. Ein Abschnitt (Paragraph) soll nur erlaubte Zeichen(ketten) enthalten.

<!ELEMENT p (#CDATA) >

Außer den "CDATA",⁸⁴ die einen definierten Zeichenvorrat verwenden, läßt SGML noch Daten zu, die von dem "Parser" nicht verarbeitet werden können, kurz die Non-SGML Daten.

"Non-SGML data is data that is not parsable in accordance with this Standard. It is either data in an undefined character set, bit data, or some mix of the two. In undefined character set data, the bit combinations represent characters, but not in the document's character set. In bit data, the bit combinations do not represent characters at all" /ISO/DIS 8879, 1985, 94/.

⁸³ Verschiedene Techniken, Elementdeklarationen zusammenzufassen und somit zu vereinfachen, sind erlaubt; vgl. /ISO/DIS 8879, 1985, 75f/.

⁸⁴ "In character data, each bit combination represents a character in the document's character set" /ISO/DIS 8879, 1985, 93/.

Das Modell kann noch durch Attribuierungen weiter spezifiziert werden. Das Modell, wie ich es bislang beschrieben habe, bestimmt die Dokumentstruktur auf Ebene der Elementkennungen. Die Elementkennungen können durch Attribute weiter spezifiziert werden. Die Attribute bestehen aus Namen, einem Zuordnungszeichen und dem Wert, der zugeordnet wird. Der Wert muß in Hochkommata eingeschlossen werden.

Die Attribute und ihre zulässigen Ausprägungen werden im Rahmen der Elementdeklarationen festgelegt.

Die Attributdefinitionen folgen dem Inhaltsmodell des Elements, dem sie zugehören. Es wird außer dem Attributnamen und dem gültigen Wertebereich noch ein Standardwert definiert, der eingesetzt wird, wenn kein anderer Wert angegeben wird (Default-Wert). In die Darstellung können noch Kommentare, durch je zwei Bindestriche an Anfang und Ende des Kommentars abgegrenzt, in die Definition eingefügt werden.

Das Element "Arbeit" könnte beispielsweise ein Attribut für die Zugangsberechtigung, für den Status und die Version enthalten.

```
<!ELEMENT --          CONTENT          --

      Arbeit          (Vorspann, Corpus, Nachspann)

--          ATTRIBUT          WERT          DEFAULT  --

          Status          (Abgabe | Entwurf)          'Entwurf'

          Zugang          CDATA          REQUIRED

          Version          NUMBER          '01'          >
```

Am Textanfang eines Dokuments "Arbeit" stünde als erste Marke möglicherweise:

```
<Arbeit Status='Abgabe' Zugang='KB' >
```

Automatisch würde dann angenommen, daß es sich um die Version 01 handelt; entgegen dem Standardwert wird bestimmt, daß es sich um eine Abgabeverision handelt.⁸⁵

⁸⁵ Auch hier sind weitere Möglichkeiten der Komprimierung und Zusammenfassung von Deklarationen außer acht gelassen worden.

Kurz gesagt sei noch, daß es ein Attribut, den "unique identifier" (ID) gibt. Seine Funktion ist, die Elementkennungen, die für alle Elemente einer Klasse benutzt werden, zu "individualisieren", so daß ein Element eindeutig ansprechbar wird. Der Sinn liegt darin, daß solche Elemente referenzierbar sind. Wird z.B. jede Kapitelüberschrift mit einer "ID" versehen, kann aus dem laufenden Text auf andere Kapitel verwiesen werden. Wie die Referenz abgearbeitet wird, liegt bei den konkreten Verarbeitungsschritten. Beispielsweise kann eine Referenz dazu führen, daß an der Stelle ein Text "Siehe dazu Kapitel 2.3 S. 45ff" erzeugt wird. Dabei enthielte die Referenz nur die ID des Kapitels 2.3. Zusätzlicher Text und Seitenangabe würden von der Verarbeitungsseite hinzugefügt.

Einen weiteren Bestandteil der Markupdeklarationen bezieht sich auf Entitäten. Ihre Bedeutung liegt darin, wie gesagt, Teile außerhalb des eigentlichen Dokumentflusses zu deklarieren. Damit ist es möglich, einen längeren Text in viele kleinere Abschnitte zu zerlegen, was gerade für die Handhabung von Texten am Bildschirm wichtig ist, aber auch von manchen Systemen gefordert werden muß, die nur Texte einer bestimmten Länge "am Stück" bearbeiten können.

Über die Entitäten können aber auch Zeichenfolgen für nicht darstellbare Zeichen definiert werden, die dann bei mächtigeren Ausgabegeräten durch die entsprechenden Zeichen ersetzt werden. Das griechische Alphabet Zeichen für Zeichen zu deklarieren wäre als Anwendung denkbar. Auch die Erzeugung deutscher Umlaute mit amerikanischen Tastaturen ginge auf diesem Weg. Diese Technik kann auch angewendet werden, um immer wiederkehrende Ausdrücke zu ersetzen. So wäre es in diesem Kapitel wünschenswert, nicht immer "Standard Generalized Markup Language" ausschreiben zu müssen. Für den Zweck kann eine Entität deklariert werden:

```
<!ENTITY SGML 'Standard Generalized Markup Language'>
```

Taucht jetzt im Text nicht einfach SGML, sondern &SGML; auf, würde bei der Dokumentverarbeitung der lange Ausdruck eingesetzt.

Es können externe Entitäten deklariert werden, die z.B. die Daten von Abbildungen als bitmap enthalten.

Es können "public entities" definiert werden, d.h. Entitäten, die z.B. Dokumenttypdefinitionen enthalten, die von einer Nutzergruppe gemeinsam verwendet werden können. Es ist selbstverständlich, daß die Deklarationen nicht jedesmal neu "erfunden" oder auch nur "getippt" werden müssen.

4.1.3.4 Markup Minimierung

Markup Minimierung wird als optionales "feature" der SGML angeboten. Ziel der Markup Minimierung ist, den Eingabeaufwand für das Markup, die Anzahl der notwendigen Anschläge zu reduzieren. Diese Minimierungstechniken laufen unter den Namen: OMITTAG, SHORTTAG, SHORTREF und DATATAG.

OMITTAG erlaubt, Markup "auszulassen", wenn in zwei aufeinanderfolgenden Elementen das zweite nicht dem ersten hierarchisch untergeordnet ist. In dem Fall versteht es sich sozusagen, daß mit dem Anfang des zweiten Elements das erste Element beendet ist. Auf eine Elementendemarke kann in dem Fall verzichtet werden.

SHORTTAG reduziert das Markup, indem Teile der Anfang- oder Endemarken von Elementen weggelassen werden können. So kann, wenn es im Kontext unmißverständlich ist, eine Endemarke für einen Paragraphen "</p>" auf "</>" reduziert werden.

DATATAG bedeutet die Möglichkeit, Daten, d.h. normale Textteile gleichzeitig als Markup zu interpretieren. Dadurch kann ein Text im Prinzip weitgehend von Markup befreit werden. So können beispielsweise Anführungsstriche als Beginn und Ende eines Zitats interpretiert werden. SHORTREF ist eine weitere Möglichkeit, Markup zu minimieren und den Text davon zu befreien.

"The trick is to define markers as entities and then to 'map' or 'associate' the entity names to a short code consisting of one or two characters." /OOPEP 1985, 15/

Besonders vereinfacht wird der Markup Prozeß, wenn statt der Ersetzung einer längeren durch eine kürzere Eingabe die längere Eingabe durch ein Steuerzeichen ersetzt wird. Das wäre z.B. der Fall, wenn man ein "Carriage Return/Linefeed" als die Anfangsmarke eines neuen Absatzes interpretieren ließe. Dadurch, daß Markup an Textverarbeitungsfunktionen angebunden wird, verschwindet es erstens aus dem Text und wird zweitens automatisch, nebenbei, erzeugt.⁸⁶

Die Markupminimierung erfordert entsprechende Markup Deklarationen, die hier nicht behandelt werden sollen.

⁸⁶ FORMEX /Vgl. Guittet 1986/ ist eine SGML-Implementation, die dieses Prinzip extensiv und "benutzerfreundlich" einsetzt.

4.1.3.5 SGML-Implementation und -Parser

Wenn man einmal eine SGML-Implementation Schritt für Schritt verfolgte, sähe das Vorgehen etwa so aus:

Die Implementation begänne bei der Analyse der Dokumente, die in Zukunft SGML-Dokumente sein werden. Das erste Ziel ist, eine Dokumenttypdefinition zu entwickeln. Dazu können die Systementwickler mit den Auftraggebern gemischte Gruppen bilden, um die für einen Anwender wichtigen Elemente einer Dokumentstruktur herauszuarbeiten. Für den einen mag eine Postleitzahl als Element wichtig sein, für einen anderen eine ISBN-Nummer. Je nachdem, ob die Dokumente auch in Retrievalsysteme eingelesen werden sollen, wären von Anfang an auch dafür Elemente zu bestimmen. Zur Erstellung der Dokumenttypdefinition werden natürlich auch Beispielsammlungen herangezogen und eventuell vorhandene Richtlinien für die Dokumenterstellung, wie sie nicht nur bei Verlagen üblich sind, sondern auch bei Institutionen, die mit Technischer Dokumentation zu tun haben.⁸⁷ Wenn die Dokumenttypdefinition provisorisch steht, werden nach ihr testweise Dokumente ausgezeichnet. An dieser Stelle kann man die Dokumenttypdefinition noch leicht ändern. Danach beginnt man ggf., Minimierungen des Markup einzuführen. Danach wird die Dokumenttypdefinition softwaremäßig getestet, d.h. sie wird in einen Parser eingelesen, der ihre Konformität mit der SGML "checkt". Wenn man sicher ist, daß die Dokumenttypdefinition der SGML entspricht, kann der Parser für SGML-Dokumente des bestimmten Typs programmiert werden.⁸⁸

Ein SGML-Parser, der, wie ein Compiler die korrekte Programmerstellung die korrekte Dokumentauszeichnung überprüft, hätte folgende Aufgaben zu erledigen.

1. Suchen und Unterscheiden von Markup (drei Varianten) im und vom Text
2. Referenzen auf Entitäten durch die Entitäten ersetzen
3. Suchen und Unterscheiden von Markup in den Entities
4. Überprüfen des Deklarationsteils auf Konsistenz

⁸⁷ Das DIN z.B. hat in einer eigenen Norm, DIN 820, festgelegt, wie DIN-Normen auszusehen haben.

⁸⁸ Dieser Ablauf ist ähnlich bei /Kennedey 1986/ beschrieben.

5. Überprüfen des deskriptiven Markup entsprechend den Anforderungen der Dokumenttypdefinition auf Gültigkeit der Kennungen, Attribute, Werte und Positionen in der Struktur.
6. Abgabe der Kontrolle an das System, um das an den Kennungen "hängende" Verarbeitungsprocedere auszuführen.⁸⁹

4.1.3.6 Zwischenbetrachtung

Hier wurden natürlich nur grundlegende Züge der SGML vorgestellt, um einen realistischen Eindruck davon zu vermitteln, welcher Art diese Sprache ist und welche Mächtigkeit sie besitzt.

Dazu zusammenfassend, bevor ich zu konkreten Anwendungen im Bereich des Elektronischen Publizierens komme:

1. Für die Erstellung von SGML-Dokumenten gibt es praktisch keine Einschränkungen. Es ist keine besondere Software erforderlich, sondern ein ganz simples Texterfassungssystem. Im Prinzip kommt man für alle Zwecke mit dem ASCII-Zeichensatz aus. Wenn ein Autor einmal nicht weiß, welches Markup einzustreuen ist, kann er, mit zwei Bindestrichen abgegrenzt, Kommentare in sein Dokument einfügen, die von einem Parser nicht verarbeitet werden, aber einem Hersteller im Verlag gelten können.
2. Ein SGML-Dokument kann physisch und logisch aus verschiedenen Dokumentteilen bestehen, ein SGML-Dokument kann Datentypen wie Rastergrafik oder Fotos integrieren, es kann Verweisstrukturen aufbauen und damit u.a. die Eingabe langer Zeichenketten durch einige Anschläge bewirken. Textverarbeitungsbefehle können als Markup "uminterpretiert" werden. Ein Dokument kann verschiedene Dokumenttypdefinitionen gleichzeitig haben; es ist möglich, Vorschriften zu formulieren, wie ein Dokument einer Dokumenttypdefinition in ein Dokument eines anderen Typs umgewandelt werden kann.
3. Die Dokumenttypdefinitionen können in Software umgesetzt werden, die die Erfassung unterstützt und die Syntax des Markup kontrolliert.
4. Ein Dokument im SGML-Format ist geräte-, system- und anwendungsunabhängig. Damit ist es flexibel für Mehrfachnutzungen einsetzbar; die Datenbasis bleibt unberührt von Hard- und

⁸⁹ Vgl. /ISO/DIS 8879, 1985, 70/

Softwareveränderungen beim Anwender; SGML-Dokumente eignen sich besonders für den Austausch von Dokumenten.

4.2 ANWENDUNGEN DER SGML

Die Markup '86 oder im Langtext "Fourth International Conference on Electronic Manuscript Preparation, The Standard Generalized Markup Language & Technical Documentation Issues" bot Gelegenheit, sich über den aktuellen Stand der Anwendungen und geplante Anwendungen der Markup-Sprache zu informieren.

Die Markup-Konferenzen werden seit 1982 von der GCA, von der schon im Zusammenhang des GenCode die Rede war, veranstaltet. Sie können als ein Forum *verschiedener* Zweige der Informationsindustrie angesehen werden. An der Zusammensetzung der Konferenzteilnehmer läßt sich recht gut ablesen, daß das Elektronische Publizieren als ein kooperatives Unternehmen im Kontext des informations-industriellen Komplexes anzusehen ist. Auf der Markup '86 waren sowohl renommierte Computerhersteller, Firmen aus dem Druck- und Satzbereich, große Verlagshäuser wie Elsevier, Springer (Heidelberg), Butterworth, Wolters Samsom Group, Larousse vertreten wie auch Anwender und Entwickler aus dem technisch-wissenschaftliche Bereich.

Obwohl die ISO/DIS 8879 erstmals September 1985 veröffentlicht wurde, ist sie doch schon wesentlich länger im Gespräch und hat auch dadurch eine gewisse Stabilität erlangt, daß mit der DCFGML von IBM ein Paradigma vorgegeben war. Schon vor der letzten Abstimmung über die SGML gibt es eine "Users' Group" von etwa 200 Mitgliedern.

Die Frage der SGML-Anwendungen geht in zwei Richtungen. Einmal ist die Frage, welche Unterstützung der Standard durch Softwareentwicklungen, konkret von Parsern und "intelligenten" Editoren, erfahren wird und zweitens, welche Unterstützung der Standard durch die Promotion von Autoren- und Verlegerrichtlinien erhalten wird. Da der Standard ja bewußt *keine* Dokumenttypen definiert, in dieser Hinsicht also flexibler als ODA ist, entsteht am anderen Ende eine zusätzliche Notwendigkeit a) die Dokumenttypen selbst zu entwickeln und b) die entwickelten Dokumenttypen und die daraus folgenden Richtlinien zur Auszeichnung entsprechender Dokumente öffentlich zu machen und für ihre Akzeptanz zu sorgen.

Im folgenden werde ich in einem Abschnitt von Arbeiten zur Entwicklung von Autoren- bzw. Verlegerrichtlinien berichten, in dem darauf folgenden Abschnitt auf Softwareentwicklungen zur SGML eingehen und in dem anschließenden Kapitel die Darstellung der Standard

Generalized Markup Language mit Überlegungen zu den Akzeptanzbedingungen der SGML abschließen.

4.2.1 Dokumenttypen und Auszeichnungsrichtlinien

Die ISO/DIS 8879, die Produktionsregeln zur sprachlichen Beschreibung von Dokumentstrukturen definiert, läßt Anwendungen und Verarbeitungsmodalitäten offen. Damit diese Offenheit nicht zur Beliebigkeit wird - und dann den freien Dokumentenaustausch behindert - sind "publikumswirksame" Aktivitäten von Verbänden und Institutionen nötig, die konkrete Dokumenttypdefinitionen entwickeln, ausführlich dokumentieren und für die Nutzer (in erster Linie Autoren, Verlage, Softwarehersteller) aufbereiten.

4.2.1.1 Das Electronic Manuscript Project

Das "Urprojekt" in diesem Sinn ist das "Electronic Manuscript Project" der AAP (Association of American Publishers), das 1983 begann und im Juli 1986 zum Abschluß kam. Ergebnisse sind der sog. AAP Standard (die Normung der Norm) und die zugehörigen Richtlinien.⁹⁰ Die ursprüngliche Zielsetzung war "to help all those involved in publishing and disseminating information to realize the greatest possible benefit from computer technology" /Sperling Martin 1985, 13-4/. Der Weg dorthin geht über das neutral ausgezeichnete Manuskript:

"Ultimately, this will be achieved when every version and derivate of a published work - whether a magazine article, a database file or a bibliographic record - can be generated directly from the original electronic form of the manuscript." /Sperling Martin 1985, 13-4/

Außer von der AAP wurde das Projekt noch vom "Council of Library Resources" gefördert. Wesentliche Projektteile wurden der Aspen Systems Corporation übertragen. In einer ersten Phase führte das Aspen Institute, eine Organisation der Wolters Samsom Gruppe, eine Befragung unter Autoren, Verlegern und in verschiedenen anderen Professionen durch, um Bedarf, gängige Praxis und Kompatibilitä-

⁹⁰ Die Ergebnisse sind publiziert. Siehe /Association American Publishers 1986 a und b/.

tätsprobleme zu erheben. Auch hier findet sich die Einschätzung, daß die Dynamik von den Autoren ausgeht:

"Authors have been promoting the electronic processing of manuscripts more extensively than have publishers" /ebd. 13-9/.

1983 wurden von den Elektronischen Manuskripten, die von Verlagen angenommen wurden, 40% als Input für die weitere Produktion verwendet.

1983 erstellten bereits 60% der befragten Autoren elektronische Manuskripte /Vgl. ebd. 13-10/.

Für die Entwicklung der Dokumenttypdefinitionen gab es eine "bibliographic task force", deren Arbeit von der National Library of Medicine beobachtet wurde.

Der AAP-Standard wurde auf Grundlage der SGML entwickelt. Für die Richtlinien sollte gelten, daß ein Anwender nur das kennen muß, was er auch wirklich braucht.⁹¹ Dementsprechend wurden Dokumenttypdefinitionen für Monographien & Bücher, Serien, Proceedings, Technische Berichte und maschinenlesbare Datenbasen erarbeitet. Die Auszeichnungsrichtlinien wurden noch weiter nach Textelementen modular organisiert in: Fließtext, der nur Zeichen nach ISO 646 enthält, solcher, der noch andere Zeichen und Symbole enthält, Tabellenmaterial und schließlich mathematische und chemische Formeln.

Der Standard ist aber nicht unumstritten. Auf der Markup '86 trug Anders Berglund einen Vergleich verschiedener Formelnotationen vor, xROFF, EQN, TeX und nach den AAP-Richtlinien. Er kam zu dem Ergebnis, daß die AAP-Notation, wenn man sie ohne Minimierungsfeatures benutzt, umständlich ist, wenn man aber die Minimierungstechniken anwendet, die Formelnotationen für den Menschen nicht mehr eindeutig zu interpretieren sind.

Ein zweiter Kritikpunkt richtet sich gegen die Beschränkung auf den ASCII-Zeichensatz (ISO 646). Die Beschränkung bedeutet, daß *alle* weiteren Zeichen als Zeichenfolgen dieses Zeichensatzes codiert werden müssen, obwohl die ISO selbst mit ISO 6937 und ISO 2022 erweiterte Zeichensätze normiert hat. Die britische Normorganisa-

⁹¹ "The standard must be so organized that authors need to become familiar only with the elements they are most likely to use in a particular manuscript. To that end the standard must consist of logical subsets of manuscript elements related to the type of document and nature of information content." /Sperling Martin 1985, 13-16/

tion (BSI) hat in der ISO erfolgreich darauf hingewirkt, daß die SGML nun auch diese Zeichensätze (und Erweiterungstechniken) zuläßt, der AAP-Standard aber hat nicht nachgezogen.

"It would be wrong if the AAP-recommended conventions did not take cognisance of this, when it is permitted in SGML" /Smith 1986 e, 198/.

Kritik an einem Standard, auch dem AAP-Standard, ist sicherlich höchst willkommen, solange er noch in der Entwicklung ist. Danach allerdings indiziert Kritik Mängel der Norm, die, wenn sie gravierend sind, die Akzeptanz der Norm beeinträchtigen. Die Kritiken, wie sie gegen den AAP-Standard vorgebracht werden, bedeuten jedoch nicht, daß dieser Standard nicht der SGML entspräche, sondern lediglich, daß die Möglichkeiten der SGML nicht "geschickt" oder nicht "hinreichend" umgesetzt werden.

4.2.1.2 STRUKTEXT

Vom Electronic Manuscript Project der AAP führt ein direkter Weg zum deutschen "Electronic Manuscript Project".

"Der Vorstand des Börsenvereins hat auf seiner 239. Sitzung den Betrag von 5000 Dollar bewilligt, um die Nutzungs- und Auswertungsrechte an dem vom amerikanischen Verlegerverband entwickelten 'Electronic Manuscript Project' zu erwerben. ... Mit dem Erwerb der umfangreichen Arbeitsunterlagen ist die Erlaubnis verbunden, daraus einen eigenen Standard für eine deutsche Auszeichnungssprache zu entwickeln." /Standard 1986, 588/

Im März 1985 war eine gemeinsame Arbeitsgruppe des Börsenvereins und des Bundesverbandes Druck geschaffen worden, um einen "Standard für strukturelle Texterstellung" /Standard 1985, 3280/ zu erarbeiten.

Interessant ist, daß diese Initiative, anders als das amerikanische Projekt, vom Druckbereich mitgetragen wird. Der Druckbereich, genauer der Satzbereich, hatte sich schon früher mit dem Problem der Inkompatibilität von Satzbefehlssprachen befaßt.⁹² Eine einheitliche Satzbefehlssprache, wie vorgeschlagen, konnte sich aber nicht durchsetzen. Aufgrund der Entwicklung der Informationstechnologien stellt sich die Frage heute auch nicht mehr allein branchenintern.

⁹² Einige Studien dazu werden in /Meinecke 1986 a/ angeführt.

Die Mikrocomputer an den Arbeitsplätzen von Autoren und in den Büros machen es zu einem vitalen Interesse der Druckbetriebe, sich für die Weiterverarbeitung der dort anfallenden Daten offenzuhalten.

Die Lösung wird in einer systemneutralen, logischen Auszeichnungssprache gesehen. Der Argumentationsgang von Meinecke /Meinecke 1986 b/ für das Engagement des Druckverbandes leitet sich weiter etwa so ab:

"Die logische Struktur ist im wesentlichen Sache des Autors. ... Aufgabe der Satzherstellung ist ... die Zusammenführung von Logikstruktur und Layoutstruktur zu einem gestalteten Dokument, einem Satzprodukt. Um generell anwendbar zu sein, sind entsprechende Auszeichnungsempfehlungen bereits für Autoren zu schaffen. Die Erarbeitung einer derartigen übergreifenden Vereinbarung ist Aufgabe der Druckindustrie und ihrer Kunden (besonders der Verlage und Autoren)." /Meinecke 1986 b, 3/

Das Projekt hat drei aufeinander aufbauende Zielsetzungen:

1. *inhaltsbezogene* Dokumentauszeichnung
2. *inhaltsbezogene* und *dokumenttypenspezifische* Dokument-Auszeichnung
3. *inhaltsbezogene*, *dokumenttypenspezifische* und *automatisationsgerechte* Dokument-Auszeichnung /Vgl. Meinecke 1986b, 4/

Mit der Vorlage der Autorensprache STRUKTEXT /siehe Krüger 1986 c/ ist die erste Phase abgeschlossen worden. STRUKTEXT ist eine Anwendung der SGML. Insgesamt ist der Ansatz "bescheidener" als das amerikanische Projekt. Während der amerikanische Standard dazu geeignet ist, Autorenrichtlinien eines Verlages zu ersetzen, ist STRUKTEXT eher als ein "Minimum" an Auszeichnung zu sehen. Daraus folgt, daß diese Sprache schnell zu erlernen ist, was die Verbreitung fördern mag. Damit ist von Anfang an auch eine Teilung der Auszeichnungstätigkeit zwischen Autor und Hersteller im Verlag für Manuskripte, die die gestattete Komplexität übersteigen, nahegelegt.

Auf den "englischen Weg", der eine dritte Strategie verfolgt, SGML-Anwendungen zu fördern, will ich nur in einer Fußnote aufmerksam machen.⁹³

⁹³ Der englische Weg hängt sehr eng mit dem "SGML Starter Set" zusammen (siehe /Smith, Stuteley 1986/ und /Smith 1986 b/). Joan Smith, die als Mitglied des BSI im ISO/TC97/SC18/WG8 aktiv

4.2.1.3 DAPHNE

DAPHNE steht für *Document Application Processing in a Heterogeneous Network Environment*. DAPHNE ist eine Entwicklung des Hahn-Meitner-Instituts (Berlin) für das Deutsche Forschungsnetz. Für die Definition der DFN-einheitlichen Dokumenttypen wird die SGML eingesetzt.

Die Dokumentverarbeitung steht im Rahmen des DFN im Kontext von Graphik- und Modellierdiensten.⁹⁴ In dem Gesamt der Dienste zur graphischen Datenverarbeitung ist die Übertragung von Dokumenten nur eine Teilaufgabe. Insgesamt lassen sich drei Komplexe bezeichnen, auf die sich die DFN-Dienste beziehen:

- "Entwurf und Realisierung eines *GKS-orientierten graphischen Dialogs* und eines *graphischen Filetransfers*
- *Übertragung und Bearbeitung von Modellierdaten* im DFN und
- Entwicklung einer Datenstruktur zur Integration von Text und Graphik in Dokumenten und *Übertragung der Dokumente* im DFN"/Maiß 1985, 3/.

Die Dokumentverarbeitung und der Dokumentenaustausch im DFN bezieht komplexe Elemente wie Formeln, Raster- und Vektorgraphiken mit ein. "Der Schwerpunkt der angestrebten Dokumentendienste im DFN wird auf

mitarbeitet, Präsidentin der SGML Users' Group ist, hat auch Autoren- und Verlegerrichtlinien im Auftrag der British Library erstellt, die demnächst erscheinen werden. Am Rande vermerkt sei, daß die nationalen Wege sich auch auffällig durch die Projektträger der SGML-Anwendungen unterscheiden. Das "SGML-Starter Set" ist bewußt stark an das "Starter Set" der IBM für die DCFGML angelehnt, verändert insofern, als es an der konkreten Syntax (reference concrete syntax) der SGML ansetzt und spezifische Elemente für den Publikationsbereich wie Copyright oder ISBN hinzufügt. Insgesamt liegt seine Komplexität erheblich über der von STRUKTEXT. Umstritten war auf der Markup '86, ob dieser Weg potentiellen SGML-Anwendern den Einstieg erleichtert oder dazu führt, daß Anwender darauf verzichten, spezifisch auf sie zugeschnittene Anwendungen zu entwickeln.

⁹⁴ Zur Begründung dieser Dienstleistungsangebote: "Bei der rechnergestützten Entwicklung komplexer technischer Produkte wächst der Bedarf, auf die aktuellen anwendungsbezogenen graphischen Informationen über spezielle Dienste eines Rechnernetzes zuzugreifen." /Maiß 1985, 1/

der dezentralen Erstellung und Bearbeitung von Dokumenten liegen. Weiterhin sollen Dokumente über das DFN ausgetauscht und weiterverarbeitet werden können" /ebd. 9/. Das bedeutet, daß einheitliche Datenformate festzulegen sind und mit Blick auf die Weiterverarbeitung auch Schnittstellen, wie sie in dem allgemeinen Dokumentmodell (vgl. "Ein allgemeines Modell der Dokumenterstellung" auf Seite 47) gefordert wurden. Die graphischen und die alphanumerischen Informationen sollen in getrennten Dateien, einer Textdatei und einem graphischen Metafile (Computer Graphics Metafile oder "GKS-Metafile") gespeichert werden. Zur Zeit ist das Datenformat für die *unformatierten* Dokumente neutral, SGML konform, für die Formatierung der Dokumente wird TeX eingesetzt, das "device independent" (dvi-files), also geräteunabhängige Ausgabedateien erzeugt. Fernziel ist, beide Dateien in einem Document Metafile (DOCM) zu integrieren mit einem einheitlichen geräteunabhängigen Datenformat.

Das Dokumentverarbeitungssystem DAPHNE kann als eine "echte" und durchdachte SGML-Implementation angesehen werden. Im Rahmen von DAPHNE sind bislang fünf Dokumenttypen definiert worden:

- report (Bericht)
- subrep (Teilbericht)
- paper (Beitrag)
- letter (Brief)
- slideset (Folien)

Zur Veranschaulichung habe ich einmal die Dokumentstruktur des Dokuments "letter" (unwesentlich vereinfacht und ins Deutsche übersetzt) aufgegriffen. In der Benutzeranleitung /Scheller, Smith 1986/ werden die zulässigen Elemente mit Hilfe von Graphen definiert. Diese Darstellungsform halte ich im Vergleich zu den Strukturbäumen für überlegen. Sie ist gewissermaßen dynamischer, da sie Reihenfolge *und* Häufigkeit des Auftretens der Elemente anschaulich machen kann.⁹⁵

⁹⁵ Die Symbolik mußte in den folgenden Abbildungen Tribut an die eingeschränkten Möglichkeiten der Textverarbeitung zahlen und entspricht nicht im strengen Sinne der Graphendarstellung. Dennoch glaube ich, daß ihre Klarheit erhalten bleibt.

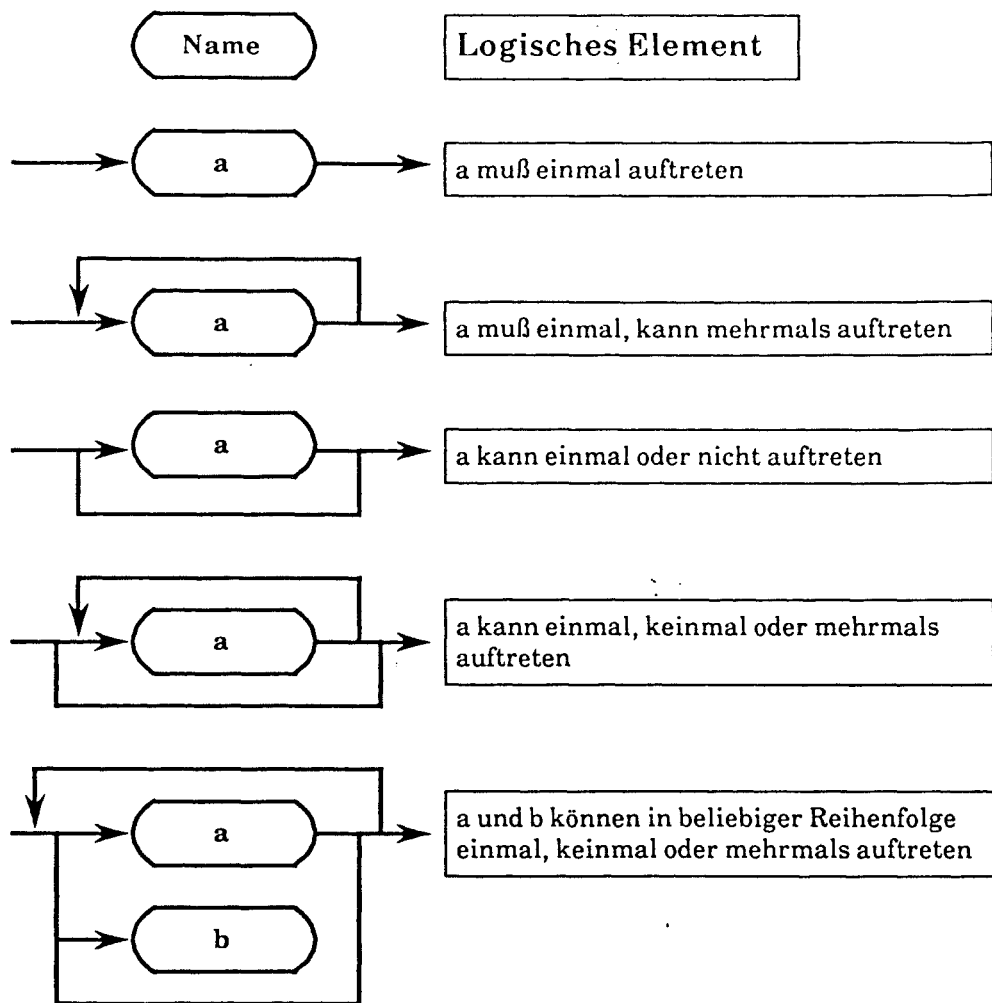
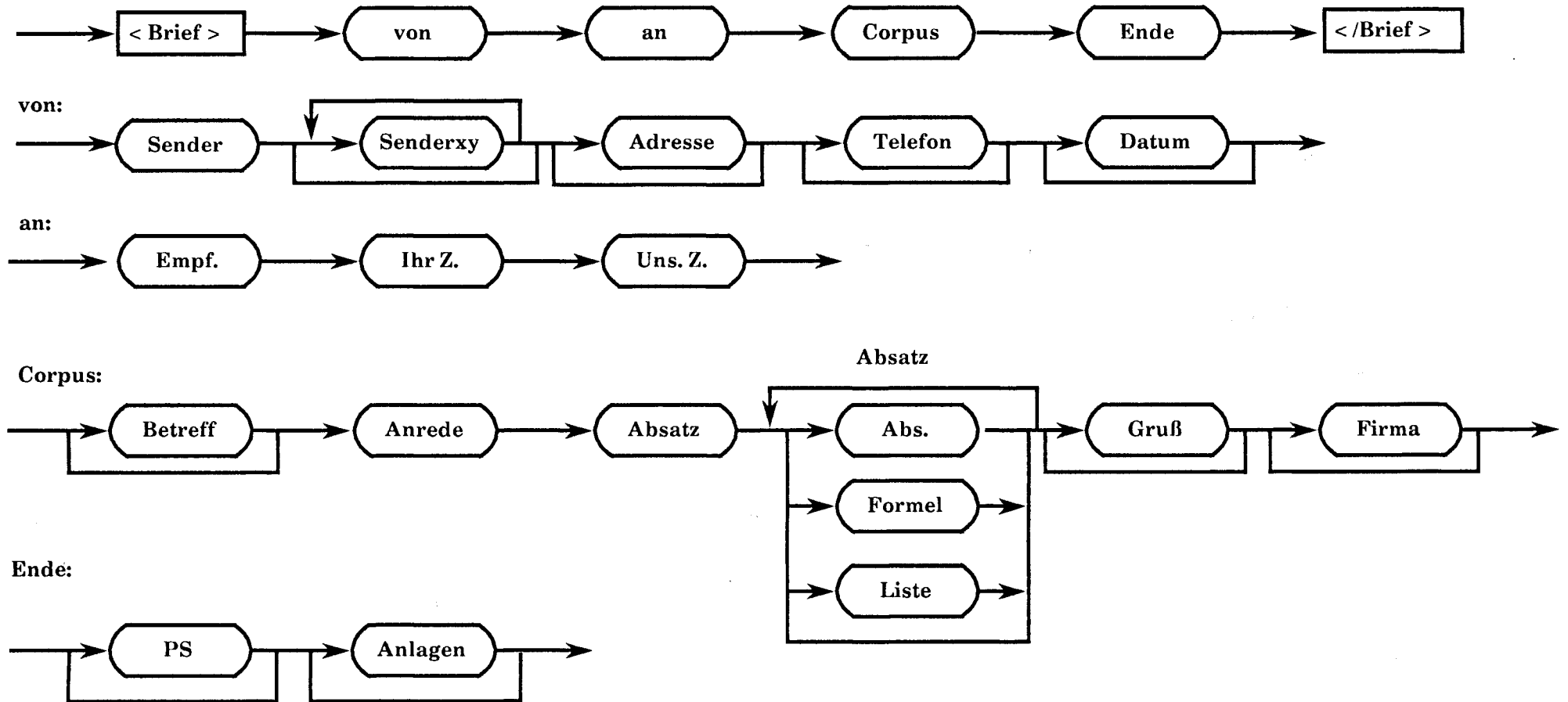


Abbildung 9. Symbole der Graphendarstellung

Die Struktur des Dokuments "Brief" besteht praktisch aus einem Anfangs- und einem Endpunkt und vier Pfaden mit jeweils spezifischen "Durchlaufmodalitäten".

Die Arbeit mit DAPHNE sieht etwa so aus, daß ein Autor seinen Text nach der Benutzeranleitung in einem beliebigen Editor erstellt, d.h. mit dem für die Dokumenttypen erforderlichen Markup versieht. Danach hat der Autor in einem Prädialog, in dem er dem System zu-

Dokumenttyp "Brief"



nächst mitteilt, um welchen Dokumenttyp es sich handeln soll (z.B. "letter"), die Möglichkeit, Einfluß auf die Gestaltung des Textes zu nehmen (falls er die Voreinstellungen ändern will). Danach wird das unformatierte, aber ausgezeichnete Dokument von einem Parser auf korrekte Syntax kontrolliert und wenn erforderlich werden Fehlermeldungen ausgegeben. Anschließend wird das Dokument formatiert und auf einem Laserdrucker ausgegeben.

An DAPHNE ist nicht nur hervorhebenswert, daß es die wohl erste öffentliche SGML-Implementation in der Bundesrepublik ist. Der Ansatz ist auch im Rahmen des Elektronischen Publizierens bedenkenswert. DAPHNE ist Teil der wissenschaftlich-technischen Kommunikation und selbst nochmals in eine Reihe von Diensten eingebettet. DAPHNE setzt bei der informellen Kommunikation an, bzw. an einem integrierten Arbeitszusammenhang. Die Anwendung von SGML im DFN greift sozusagen in die wissenschaftlichen Sozialisationsbedingungen und Arbeitszusammenhänge ein, und d.h.: logische Textauszeichnung steht nicht erst dann ins Haus, wenn es ans Publizieren geht. Wenn dann aber veröffentlicht werden soll, kann das DFN praktisch einen Ausgang an die publizistische Öffentlichkeit bieten. Und tatsächlich werden zur Zeit Überlegungen angestellt, ob STRUKTEXT nicht der sechste Dokumenttyp des DFN werden soll.

4.2.2 Anwendersoftware zur SGML

Für die Arbeit mit SGML - vom Autorenstandpunkt wie vom Verlegerstandpunkt aus - ist Software zu fordern, die die Eingabe des Markup erleichtert. In der ISO, um das nochmal aufzugreifen, ist das Thema Gegenstand einer eigenständigen Norm zu den sogenannten SGML-smart-editors und eines technischen Berichtes zum Umgang mit SGML auf konventionellen, nicht dezidierten SGML-Editoren. Bislang gibt es aber noch keine veröffentlichten Entwürfe zu den beiden Vorhaben.

Unterstützung der SGML durch Programme (Software) kann sich auf den Erfassungsaufwand richten, gemessen in Anschlägen oder Zeitdauer, sie mag aber auch danach bemessen werden, wie sie Fehler im Markup meldet und die Fehlerbehebung unterstützt oder auf andere Aspekte "komfortabler Benutzeroberflächen".

4.2.2.1 Markup mit einfachen Textverarbeitungssystemen

Die Softwareunterstützung kann im einfachsten Fall bereits da beginnen, wo logische Auszeichnungen auf Funktionstasten gelegt wer-

den. Eine Vereinfachung des Markup kann aber auf vielen Textverarbeitungssystemen noch weiter getrieben werden. Ein Benutzer kann die Eingabearbeit weiter reduzieren, wenn er lange Markierungen (Elementkennungen + Attribute z.B.) als Textbausteine speichert und bei Bedarf abrufen. Denkbar ist auch, daß in einem Fenster der Text, in einem anderen eine Übersicht der Kennungen oder die ganze Dokumenttypdefinition angezeigt wird. Eine weitere Möglichkeit bieten Textverarbeitungssysteme, die auf die ein oder andere Weise dem Benutzer erlauben, in den Text eingefügte SteuerCodes des Systems über Suchen+Ersetzen-Funktionen in Markup umzusetzen. Je nach Textverarbeitungssystem ergeben sich weitere Möglichkeiten, den Erfassungsaufwand herunterzuschrauben.⁹⁶ Diese Systeme, so komfortabel sie auch sind, stellen keine Implementationen von Dokumenttypdefinitionen dar, d.h. sie nutzen die Informationen über die Dokumentstruktur nicht. Sie bleiben letztlich reine Erfassungsgereäte.

4.2.2.2 Konvertierprogramme

Es ist eine einleuchtende Idee, das ganze Auszeichnungsproblem Konvertierprogrammen zu überlassen. In der Tat ist das kein ungewöhnlicher Weg. Das Aspen Institute stellte z.B. ein Konvertierprogramm von "WordStar" in den "AAP-Standard" auf der Markup '86 vor. Aber auch in der Bundesrepublik ist ein Konvertierprogramm zur Erreichung des "neutralen" Formats im Rahmen des DOCDEL-Projektes P14 "Elektronisches Publizieren technisch-wissenschaftlicher Texte" zum Einsatz gekommen. Um von Autoren angelieferte maschinenlesbare Texte für den Satz und für die Eingabe in eine Datenbank übernehmen zu können, setzte der Nomos-Verlag, Konsortium im "Verlegerkonsortium", das Konvertierprogramm KONSA ein und war erfolgreich:

"Der Weg, den der Nomos-Verlag mit dem Einsatz eines einfachen und marktgängigen Textverarbeitungssystems (wordstar) bei seinen Autoren - und der anschließenden Konvertierung für Datenbankaufbau und Fotosatz - beschritten hat, führte zum Erfolg.

Die Autoren sind mit der Einfachheit der Erfassung höchst zufrieden" /Projekt P14 1986, Seite d/.

⁹⁶ Der Eingabeaufwand hängt aber auch entscheidend davon ab, ob die Dokumenttypdefinition "Minimierung" des Markup vorsieht. Wie bereits erwähnt, läßt sich damit nicht nur die Zahl der Anschläge reduzieren, sondern auch eine Verbindung von SteuerCodes der Textverarbeitung und Markup herstellen.

Es muß allerdings vorweg gesagt werden, daß sich die Lösung auf einfache Texte bezieht, d.h. keine Graphiken und Formeln berücksichtigt werden können. Tabellen bis zu einer gewissen Komplexität werden jedoch noch konvertiert.

Das Konvertierprogramm KONSA (kundenorientierte nicht satzmaschinenorientierte Textverarbeitung) kann entweder bei dem Autor, beim Verlag oder beim Satzbetrieb vorhanden sein. In dem Projekt war die Probe auf's Exempel die Konvertierung von Beiträgen zur "Zeitschrift für öffentliche und gemeinwirtschaftliche Unternehmen" (ZögU), die mit WordStar erfaßt worden waren /Vgl. Projekt P14 1986 1-4/.

Das Verfahren beginnt bei der Erfassung des Textes mit WordStar. Für Korrektur- und Lektoratszwecke wird das Manuskript auf einem Schnelldrucker ausgegeben. Die auf dem Papier gemachten Änderungsvorschläge und Korrekturen werden am Elektronischen Manuskript mit WordStar vorgenommen.

Wenn das Manuskript druckreif ist, wird das Konvertierprogramm KONSA eingesetzt. Eine Satzanweisung muß allerdings vorliegen. Wie üblich wird eine Satzanweisung (Informationen über Schriftarten, Schriftgrade, Auszeichnungsarten, Satzspiegel etc.) für den Satzrechner benötigt. In diesem Fall ist aber noch eine weitere Tabelle erforderlich, die WordStarcodes dem zuordnet, was als Fotosatzausgabe erwünscht ist. Das Programm KONSA wird daraufhin instruiert, welche Makros einer bestimmten Photosatzsprache die WordStarcodes ersetzen sollen. Die mit KONSA konvertierte Datei ist noch nicht vollständig für ein *bestimmtes* Layout festgelegt, da beispielsweise einem Makro für Hervorhebungsstufe 1 noch verschiedene Fotosatzbefehle zugeordnet werden können (unterstreichen, kursiv, gesperrt etc.).

So elegant das Programm auch sein mag, so wird doch bereits deutlich, daß die Struktur eines Textes nicht komplexer sein darf, als es mit WordStar darzustellen ist. Für die Verwendung von WordStar für den späteren Fotosatz sind besondere Erfassungsanweisungen einzuhalten; es darf z.B. kein Blocksatz verwendet werden, die Returnntaste darf *nur* verwendet werden, um einen Absatz zu markieren etc.. Andere Regelungen betreffen Feinheiten für den Satz; z.B. die Unterscheidung von verschieden langen Querstrichen - Minuszeichen, Bindestrich ("couple"), Gedankenstrich - durch Kombinationen aus "Blanks" und Querstrichen.

Das Programm KONSA wurde ebenfalls dazu benutzt, die layoutorientierten WordStar-Codierungen in eine logische Auszeichnung umzusetzen. Dazu wurde eine einfache Dokumenttypdefinition für die ZögU entwickelt und eine Korrespondenz der logischen Elemente zu den WordStar-Codes hergestellt, die ihrerseits wieder auf Kombinationen

von alphanumerischen Zeichen und Steuerzeichen des ASCII-Zeichensatzes rückführbar sind.⁹⁷

4.2.2.3 Intelligente Editoren

Nicht erst seitdem es SGML gibt, gibt es syntaxgesteuerte Editoren, d.h. Editoren, die als Eingabe Zeichenketten verarbeiten, die gemäß einer bestimmten Grammatik strukturiert sein müssen. Solche strukturierten Objekte können auch Dokumenttypdefinitionen sein. Ein syntaxgesteuerter Editor muß grundsätzlich lexikalische Einheiten, hier: Markup und Textelemente des Dokuments erkennen und unterscheiden können und die richtige Abfolge des Markup, die Einhaltung der Grammatik, überprüfen können. Ein intelligenter Editor dieser Art ist z.B. von der Firma EPSILON entwickelt worden. Aber auch die Apollo-workstation arbeitet z.B. mit solch einem Editor. Einen "intelligenten", syntaxgesteuerten Editor kann man sich als Verwalter einer Baumstruktur vorstellen. Es können dann nochmal solche Editoren herausgehoben werden, die nicht nur *eine* Baum- oder Dokumentstruktur verwalten, sondern parametrisierbar sind, die also mit spezifischen Strukturen zu programmieren sind.⁹⁸

Ein intelligenter Editor ist aber nicht nur als Parser zu beurteilen, sondern auch als Eingabesystem für Markup, als *Textauszeichner*, wie der Terminus lautet, der im Projekt P14 favorisiert wurde.

DER INTELLIGENTE EDITOR VON DATALOGICS: Auf der Markup '86, die um eine "Software-Gallery" bereichert war, zeigte die Firma DATALOGICS eine Software, die unter MS-DOS läuft. Das ist insofern anzumerken, da bislang intelligente Editoren eher in der mittleren Datentechnik zu finden waren.

Die Software von DATALOGICS ist mit einer bestimmten Dokumenttypdefinition programmiert, die in Abstimmung mit den Anwendern ent-

⁹⁷ Bei der Einspielung der neutral ausgezeichneten elektronischen ZögU-Manuskripte in eine Datenbank auf einem Siemens-Rechner zeigte sich, daß der Rechner den ISO-7-bit-Code verarbeitet, während WordStar mit einem 8-bit-Code arbeitet. Es war dann trotz der neutralen Auszeichnung noch eine weitere Programmierung nötig (zur "Maskierung" des 8ten bits).

⁹⁸ Zu dem System *Seek* der Firma Epsilon vgl. /Projekt P14 1985 b, 129- 136/.

wickelt wurde. Sie weist mehrere bemerkenswerte Züge hinsichtlich der Benutzeroberfläche für die Textauszeichnung auf, die einen Maßstab für andere SGML-Softwarepakete abgeben.

1. Die Textauszeichnung erfolgt durch die Dokumenttypdefinition kontrolliert. Eingabefehler beim Markup können so nie zu Verarbeitungsfehlern führen. Die Fehlermeldungen werden in dem Moment gegeben, in dem der Fehler sich ergibt. Anders als bei einem "check" in einem Batch-Lauf ist so die Fehlersuche im Dokument kein eigentliches Problem mehr.
2. Funktionstasten laden sich automatisch mit Auszeichnungen, je nachdem welche Dokumentelemente in dem aktuell bearbeiteten Ausschnitt der Dokumenttypdefinition erlaubt sind.
3. Zusätzlich sind in einem Fenster die im Kontext erlaubten Marken und die Funktionstastenbelegung angezeigt, was die Auswahl der richtigen Taste absichert und beschleunigt.
4. Wenn über den Wert, den ein Attribut annehmen soll, aus einer Anzahl von Optionen zu entscheiden ist, werden die verschiedenen zur Auswahl stehenden Werte angezeigt, so daß sich der Nutzer interaktiv für einen aus dem Menü entscheiden kann.
5. Ein optional aktivierbarer Bildschirmformatierer nutzt die Fläche des Bildschirms, um logische Strukturen zu visualisieren. Einrückungen von Absätzen oder Listenelementen können z.B. wiedergegeben werden. Hervorhebungen verschiedener Stufe werden durch verschiedene Farben angezeigt bzw. durch fette, unterstrichene, inverse oder blinkende Darstellung.
6. Die Markierungen erscheinen *nicht* im ausgezeichneten Text, sondern sind vollständig (mit Attributen) auf einer zweiten Ebene bzw. in einem Fenster anzusehen und bei Bedarf auch auf dieser Ebene veränderbar

Wenn ich einmal die DCFGML zum Vergleich heranziehe, die ich ja zur Erstellung dieses Textes u.a. verwende, werden die Vorteile der beschriebenen Implementation bzw. Software noch deutlicher.

Marken im Text erschweren den Überblick über den Inhalt, um den es eigentlich gehen soll, weil sie den Text zerstückeln *und* aufblähen. Entsprechen drei Bildschirmseiten normalerweise ungefähr einer DIN-A 4-Seite Ausgabe auf Papier, kann sich dieses Verhältnis durch Markierungen im Text auf 5:1 verschlechtern.

Der Text wird auch dadurch unübersichtlich, daß Kapitelnummern, Fußnotenzählungen u.ä. erst im Formatierlauf errechnet werden. Das ist im Prinzip zwar begrüßenswert, erfordert am Bildschirm aber eine

ersatzweise Orientierungshilfe anderer Art, z.B. durch farbige Hervorhebung von Überschriften verschiedener Tiefe.

Die Fehler der Eingabe werden erst im Batchlauf entdeckt und erfordern dann oft einen aufwendigen Gang der Fehlersuche. Daß die Fehlermeldungen z.T. Kenntnisse des angeschlossenen Formatierers erfordern, die der einfache Nutzer nicht erlernt hat, darauf wurde schon an anderer Stelle hingewiesen.

Besonders gravierend ist m.E. aber, daß die Eingabe der Marken nicht unterstützt wird, was einen erheblichen Zeit- und Tippaufwand bedeutet. Vier Zeichen vor jedem logischen Element und vor jeder Hervorhebung und nochmals danach "summieren sich".

Die Entschädigung für die Mühen ist der wohlgestaltete Ausdruck auf dem Laserdrucker. Da allerdings tritt ein Problem anderer Art auf. Man kommt sich leicht "ästhetisch" bevormundet vor, oder anders: man hätte gern mehr Einfluß auf die Gestaltung. Das kann dazu führen, daß man sich doch mehr als beabsichtigt mit den Formatierbefehlen oder anderen Möglichkeiten, den "Gestaltungsspielraum" auszuweiten, befaßt. Das allerdings widerspricht der Idee eines "housestyle", der im Rahmen technischer Dokumentation oder auch bei Verlagen sehr sinnvoll sein kann.

DER TEXTAUSZEICHNER IM PROJEKT P14: Eine ähnliche Software wie der intelligente Editor von Datalogics wurde auch im Rahmen des Projekts P14 des DOCDEL-Programms der EG entwickelt. Ihr Anwendungsbereich sollte die Auszeichnung von DIN-Normen sein.⁹⁹ Die Software wurde ausgehend von einem existierenden Editor, der in CDL2 geschrieben war, entwickelt. "Der Texteditor wurde um die neuen, speziell für den Auszeichnungsprozeß benötigten Editierfunktionen erweitert" /Projekt P 14, 12/. Die zusätzlichen Funktionen und die Benutzeroberfläche wurden in C geschrieben.¹⁰⁰ Ablauffähig ist das Programm auf einem Siemens-PC-X mit dem Betriebssystem SINIX, einem UNIX-Derivat. Anders als bei geschlossenen Inhouse-Lösungen, in denen die Texte ausgezeichnet werden, *wo* sie entstehen und *beim* Entstehen,

⁹⁹ Zu den Erfahrungen mit der experimentellen SGML-Anwendung im Verlag Technische Regelwerke (VTR) sind die Veröffentlichung der Ergebnisse und der Evaluation des Projekt(teil)s durch das sozialwissenschaftliche Forschungsinstitut AGAS abzuwarten. Zu der Projektanlage von der Eingabeseite her und als "SGML-Projekt" vgl. Krüger "The Basics for Electronic Publishing of German Standards" /Krüger 1985, 158 -176/. Mehr zu der Nutzungsseite vgl. /Bengs, Mohr 1985/.

¹⁰⁰ Weitere technische Details zur Programmentwicklung finden sich ebenda, führten hier aber zu weit.

ist die Situation im DIN dadurch charakterisiert, daß die Texte auf verschiedenen Textverarbeitungssystemen andernorts erstellt werden, deren maschinenlesbares Speicherformat weiterzuverarbeiten ist. Das Programm muß von daher ein Textauszeichner *ex post* sein.

In den fertig erstellten Text - von Steuerzeichen des liefernden Textverarbeitungssystems befreit - muß das Markup "eingelassen" werden. Der Cursor wird zu dem Zweck an die Stelle im Text bewegt, an der eine Markierung angebracht werden soll. In einem Fenster werden die im Kontext (auf die Dokumenttypdefinition DIN-Norm bezogen) erlaubten Markierungsoptionen angezeigt. In dem Menü wird die gewünschte Marke angesteuert und durch Tastendruck an die Textstelle befördert, die durch die Cursorposition angezeigt wurde.

Der Markupprozeß wird in Phasen zerlegt:

- das Grobmarkup für die spezifischen Dokumentelemente
- das Feinmarkup für die Grundelemente und
- ein dritter Gang für die Auszeichnung von Tabellen.

Die Entscheidung für die Phase, die bearbeitet werden soll, wird über eine Menüauswahl dem Auszeichner mitgeteilt.

Die Aufteilung des Markupprozesses in separate Arbeitsschritte ist nicht nur vorteilhaft für die Auszeichnung von Texten im nachhinein; sie kann auch als Prinzip einer Arbeitsteilung zwischen Autor und Verlag gelten, um eine kontrollierte Zuweisung von Auszeichnungsfunktionen zu erreichen.¹⁰¹

¹⁰¹ Dieses Programm wurde später für den Springer Verlag (Heidelberg) zu einem Autorensystem, d.h. einer Auszeichnungssoftware für Autoren weiterentwickelt und ist dort unter MS-DOS ablauffähig.

5. SCHLUßBETRACHTUNGEN

Dieses abschließende Kapitel ist zweigeteilt. In dem ersten Teil wird die Erörterung der Standard Generalized Markup Language unter dem Gesichtspunkt der Akzeptanz zusammengefaßt und abgeschlossen. In dem zweiten Teil wird die äußere Klammer, um das Elektronische Publizieren geschlossen.

5.1 ÜBERLEGUNGEN ZUR AKZEPTANZ DER SGML.

Wenn die Durchsetzungschancen und -modalitäten des Elektronischen Publizierens (auch) von internationalen Standards abhängen, stellt sich die Frage nach den Akzeptanzbedingungen der Standards, hier der ISO 8879.

Lesbarkeit und Verständlichkeit ist sicherlich eine elementare Voraussetzung für die Akzeptanz eines Standard. Da es oft Klagen über die Verständlichkeit des Textes der ISO/DIS 8879 gegeben hat /vgl. Heath 1986, 7/, wurde für September 1986 eine erheblich klarere Schlußfassung angekündigt.

Die Akzeptanz der Norm kann von der ISO wesentlich durch die Entwicklung flankierender und ergänzender Maßnahmen gefördert werden, die den potentiellen Normanwendern spezifische Hilfestellungen zum Verständnis der Norm geben (z.B. terminologische Arbeiten, Beispielsammlungen, Starter Set), aber auch im Zusammenhang der SGML regelungsbedürftige Materien aufgreifen, wie die Frage der Zeichensätze, Fonts oder des Dokumentenaustausches (aufschließend zum ISO/OSI-Modell). Für die Softwarehersteller besonders wichtig wird die Norm zu den dedizierten SGML-Editoren werden und für den Verlagsbereich die Norm, die die Verfahren festlegt, wie "öffentliche" SGML-Anwendungen registriert werden können. Dieser zusätzliche Regelungsbedarf war als Konsequenz der Offenheit der Norm herausgestellt worden.

Eine weitere Möglichkeit, die Akzeptanz der Norm zu erhöhen und ihren Anwenderkreis zu vergrößern, besteht darin, sie mit der "konkurrierenden" ODA - die auch im ISO Normungsverfahren steht - zu harmonisieren. Dazu gab es im Lauf der SGML-Entwicklung Anpassungen innerhalb der Norm. Der springende Punkt ist, daß SGML ursprünglich keine Möglichkeit vorsah, auf Dokumentebene die logische Struktur mit einer Layoutstruktur zu verbinden. Die Aufnahme der "concurrent document types" und neuerlich der Vorstoß, sogenannte

"Link Type Declarations" in die Norm aufzunehmen, resultieren wahrscheinlich aus diesem Harmonisierungsbestreben. Das perfekte Ergebnis der Harmonisierung ergäbe, daß ODA eine regelrechte SGML-Implementation (für den Bürobereich) wäre.¹⁰²

"... the distinction between the two becomes more blurred as time goes on. Indeed, it is possible for an SGML document or sub-document, structured in accordance with the office document architecture (ODA), to be transferred using the office document interchange format" /Smith 1986 a, 6/.

Als gemeinsames Forum der "SGML-Initiative" existiert eine SGML Users' Group, die ihre Mitglieder über Entwicklungen rund um die SGML und -Anwendungen auf dem laufenden hält durch halbjährliche Zusammenkünfte und ein "Bulletin". Solche Gruppen sind im Computerbereich keineswegs unüblich und für die Akzeptanz der Norm nicht zu unterschätzen.¹⁰³

Außer nach den Aktivitäten, die die "Normproduzenten" zur Akzeptanz "ihrer" Norm entfalten, muß gefragt werden, was die Softwarehersteller daraus machen werden, was die Verleger und "Inhouse-Publisher". Hier liegt auf den ersten Blick ein Henne-Ei-Problem vor.

Die Softwarehersteller von Textverarbeitungsprogrammen und Editoren und die Hersteller von Satzprogrammen warten auf eine entsprechend große Nachfrage und die Verlage auf eine komfortable Software. In der Realität sieht es ein bißchen anders aus.

¹⁰² Diese Position vertrat z.B. S. Schindler von der TU Berlin, der z.Z. einen intelligenten Editor für ODA und SGML entwickelt, auf der Markup '86. Im übrigen wird die *unterstellte* Normeninkompatibilität auch auf das (Des)Interesse der Computerhersteller zurückgeführt, die stärker im Bürobereich engagiert sind und eher ODA favorisierten. Auf der anderen Seite führte IBM die "Riege" der "SGML-Interessierten" an.

¹⁰³ Nebenbei bemerkt sollte man bei der Durchsetzung einer Norm die Rolle persönlichen Engagements nicht unterschätzen. Im Bereich der SGML sind Goldfarb für die USA, Joan Smith für Großbritannien und Manfred Krüger für die Bundesrepublik - zumindest zur Öffentlichkeit hin - die aktivsten Persönlichkeiten. Wollte man einen "personalistischen" Ansatz weiterverfolgen, müßte man von der (internationalen) Users' Group zu den nationalen Normausschüssen, die sich mit SGML und ODA befassen, übergehen.

Die Vorteile der SGML sind für den Inhouse-Publisher erheblich. Es ist kein Wunder, daß die Implementationen, die bislang existieren, hauptsächlich für Zwecke der Technischen Dokumentation und für große nicht-kommerzielle Verleger im administrativen Bereich eingesetzt werden. Beispiele sind das Department of Defense in den USA, das Office for Official Publications of the European Communities, Her Majesty's Stationary Office in Großbritannien.¹⁰⁴ Bei den geschlossenen Lösungen kommen viele Vorteile der SGML besonders stark zum Tragen. Dokumenttypdefinitionen können bis ins Feinste ausgearbeitet werden. Darauf aufsetzend kann die Erfassung der Dokumente durch Techniken wie Markupminimierung und Entity-Referenzierung rationalisiert werden.¹⁰⁵ Die Kontrolle der Auszeichnungstätigkeit durch einen Parser kann sich durchaus rechnen. Als weiterer Vorteil ist der einheitliche 'housestyle' anzusehen, der quasi zwangsläufig eingehalten wird. Große Vorteile ergeben sich auch überall da, wo ein Schriftstück eine sukzessive Bearbeitung bis zur Endfassung erfährt - wie bei Normen, Gesetzesvorlagen u.ä. - und dort, wo häufige Aktualisierungen - wie bei technischer Dokumentation von Panzern, Flugzeugen, Software u.a. - dringend geboten sind. Da Dokumentationen über längere Zeiträume archiviert werden müssen, länger als Hard- und Software in einer Organisation konstant bleiben, ist ein neutrales Speicherformat dringend geraten, um nicht von veralteter Technik abhängig zu werden. Die Schnittstelle durch SGML zum Druckbereich "Außer-Haus" soll abschließend genannt werden.

So gesehen ist es folgerichtig, daß die Dynamik der konkreten SGML-Implementationen aus diesem Bereich kommt, oder andersherum, daß SGML dort zuerst akzeptiert wird.

Für den kommerziellen Verleger sollte die SGML eine Brücke zu den Autoren mit elektronischen Manuskripten schlagen und den Weg zu einer mehrfachen und "elektronischen" Nutzung seiner Daten vereinfachen. Unter dem letztgenannten Gesichtspunkt sind eigentlich nur die großen Verlage, die man auch ruhig Medienkonzerne (Beispiel Bertelsmann) nennen kann, aktiv. Dem kleinen oder mittleren Verle-

¹⁰⁴ Siehe /OOPEC 1985/, /Guittet 1986/ und /Stutely 1986/

¹⁰⁵ Stephen Zaslav von Digital Equipment Corporation /Zaslav 1986, 2/ gab ein Beispiel, daß in einem Text von 400000 Zeichen allein 29000 Zeichen für Markup und SteuerCodes eingegeben wurden. Mit Techniken der Markupminimierung und der Fehlerbehandlung konnten die Satzkosten (hauptsächlich Erfassung) von zwei Millionen Dollar um eine halbe Million (!) gesenkt werden /Zaslav 1986, 14/.

ger drängt sich eine solche Aktivität offensichtlich nicht auf.¹⁰⁶ Für die Verlage ist der "Einstieg" in SGML sicherlich mit der Übernahme von maschinenlesbaren Autorenmanuskripten verbunden, sei es aufgrund des Drängens der Autoren oder aufgrund von Rationalisierungsgesichtspunkten in den Verlagen selbst. Auf die prinzipiellen Vorteile soll an dieser Stelle nicht nochmals eingegangen werden.

Unter der Fragestellung der Akzeptanzbedingungen der SGML ist nicht nur relevant, welchen Vorteil beide Seiten aus dem neutralen Format ziehen, sondern wie die Arbeitsplätze beim Autor und im Verlag sich verändern werden und wie die Autor-Verlagsschnittstelle aussehen wird.

Vom Autorenarbeitsplatz her, scheint die Akzeptanz der SGML daran geknüpft, ob SGML die Erfassungsarbeit erleichtern oder die inhaltliche Qualität des Manuskripts verbessern helfen kann. Dieser letzte Aspekt klang bei den Autoren der Benutzeranleitung zu DAPHNE und Meinecke vom Bundesverband Druck an /Scheller, Smith 1986, 1/ und /Meinecke 1986 b, 3/. Der Autor könne sich auf die logische Struktur seines Textes "konzentrieren", oder die logische Struktur sei Sache des Autors (die Gestaltung aber nicht). Daran richtig erscheint mir, daß ein Autor selbstverständlich nicht die Auszeichnungstätigkeit eines Herstellers im Verlag aufgehalst bekommen möchte und i.d.R. mit dieser komplexen Materie auch gar nicht ausreichend vertraut ist. Andererseits möchte sich ein Autor wohl auch gerne auf die logische Struktur seines Textes konzentrieren, was aber nicht damit verwechselt werden darf, daß die Eingabe von Markierungen diese Konzentration fördere.¹⁰⁷ Das heißt aber, daß die

¹⁰⁶ Das Fachinformationsprogramm der Bundesregierung /1985/ sieht darin offensichtlich ungenutzte Chancen, denen nachgeholfen werden soll: "Es sollen deshalb Fachverlage an die Strukturierung, Erfassung und Erschließung von Texten in Zusammenhang mit dem Aufbau von Informationsbanken herangeführt sowie technisch und wirtschaftlich beraten werden"/ebd. 23/.

¹⁰⁷ Außerdem wäre genauer zu prüfen, welche Umstellung es bedeutet, von einer integriert logisch-gestalterischen Manuskripterstellung zu einer rein logischen überzugehen.

Eine weitere Frage ist - am Rande bemerkt -, ob ein logisch (generisch) ausgezeichnetes Dokument genug Informationen für die spätere Formatierung enthält. Es ist ja durchaus denkbar, daß jemand ein *logisches Element* an unterschiedlichen Stellen in seinem Dokument *verschieden formatiert* haben möchte. In dem Fall reicht eine 1:1 Umsetzung von Markup in Satzbefehle nicht

Auszeichnungstätigkeit "softwaremäßig" gestützt werden muß, wenn sie nicht zu einer Belastung führen soll. Anhand der beschriebenen Textauszeichner wurden einige Anforderungen deutlich:

- keine (möglichst wenig) Markierungen im Text
- direkte Bildschirmformatierung logischer Elemente
- Fehlerkontrolle direkt beim Eingeben des Markup
- Unterstützung der Eingabe durch "Windowtechnik", "Menütechnik", Funktionstastenbelegung

Die danach gestaltete Benutzeroberfläche könnte so als eine modifizierte WYSIWYG-Oberfläche, wie sie z.B. vom XEROX-Star bekannt ist, beschrieben werden. Zwar sollte eine gestaltungsorientierte Formatierung am Bildschirm bereits bei der Eingabe erfolgen, aber nicht soweit, daß schon Typographie und Seitenumbruch zu berücksichtigen wären. Denn diese sollen ja erst in späteren Prozessen variabel bestimmt werden können. Das Wichtige ist also nicht, daß man sieht, was man bekommen wird (What you see is what you get), sondern daß logische Gliederungspunkte eine Entsprechung durch Anordnung auf der (Bildschirm)-Fläche bekommen bzw. durch Auszeichnungen am Bildschirm visualisiert werden. Erst dann wird es auch verzichtbar, die Markierungen als Ordnungshilfe im Text selbst anzuzeigen, und sinnvoll, sie indirekt einzugeben und auf einer zweiten Ebene automatisch zu erzeugen. Eine Funktionstaste für den Beginn eines Absatzes erzeugte im Vordergrund eine Leerzeile gefolgt von einer Einrückung, während im Hintergrund die Markierung für Absatzanfang generiert würde. Für die Weiterverarbeitung wäre nur die Hintergrundebene wichtig, die das vollständig ausgezeichnete Dokument enthielte.¹⁰⁸

Da aber solch intelligente Editoren die Implementation einer Dokumenttypdefinition voraussetzen und außerdem meistens noch auf 32-Bit-Rechnern realisiert sind, werden sie eher beim Verlag als beim Autor zu finden sein. Die Entwicklung von 32-Bit-Mikrocomputern zeigt allerdings an, daß sich das in einigen Jahren schon geändert haben kann.

aus. Es besteht aber keine Möglichkeit, direkt über das Markup mitzuteilen, daß man eine Elementkennung kontextabhängig (unterschiedlich) umgesetzt haben möchte.

¹⁰⁸ Auch für dieses Prinzip hält der Appendix B der ISO/DIS 8879 einen Namen bereit: "The presence of the entity declarations in effect, turns typewriter keystrokes into generalized markup, so 'what you see isn't all you've got' (WYSIAYG)!" /ISO/DIS 8879 1985, 118/.

Für den Verlag, der sich auf SGML einläßt, ist solch ein Editor+Parser allerdings unbedingt erforderlich, da die eingehenden Manuskripte nachgearbeitet werden müssen und vor der weiteren Verarbeitung eine maschinelle Syntaxkontrolle unumgänglich ist. Wird der Syntaxcheck unterlassen, übertragen sich die Fehler auf den Formatierer und führen dann dort zu Fehlern oder unerwünschter Ausgabe. Es reicht ja bereits, eine Endemarke einer Hervorhebungsstufe 1 zu vergessen, und man bekommt z.B. einen Text von 100 Seiten unterstrichen oder kursiv gesetzt.

Die Parser kontrollieren die SGML-Dokumente auf regelgerechte Auszeichnung. Wer aber garantiert, daß ein Parser ein 100% SGML-Parser ist? Wer garantiert, daß ein Dokument, das bei *einem* Parser ohne Fehlermeldung blieb, auf jedem Parser als gültig erklärt wird, und wer garantiert, daß ein Dokument, das keine Fehlermeldungen erzeugt, auch korrekt nach SGML ausgezeichnet ist? Kurzum, ein Parser kann die Konformität von Dokumenten zum Standard nur testen, wenn er selbst der Norm entspricht. So etwas wie ein Siegel 100%-SGML wäre eine vertrauensbildende Maßnahme, die vor allem für den *Austausch* von SGML-Dokumenten wichtig ist und eine Bedingung darstellt, die Akzeptanz der SGML zu erhöhen. Jim Heath vom National Standard Office hat vorgeschlagen, Testreihen für Parser zu entwickeln, die zwar nie mit absoluter Sicherheit die Normkonformität ergeben können, aber doch die Unsicherheit verringern.

"The primary purpose of conformance testing is to establish whether the implementation being tested conforms to the specification of the standard. We expect other benefits to derive from this:

1. "... positive influence to developers of SGML parsers
2. The fact that a parser has successfully completed a standard set of tests will improve its acceptability to users and give them confidence that they may process their important documents without misinterpretation.
3. As a result of one 1 and 2 above, we expect that the acceptance of SGML as a standard will accelerate" /Heath 1986, 7/.

Ein anderer Weg, den Autoren SGML anzumuten, ist, sie weiter ihr gewohntes Textverarbeitungsprogramm benutzen zu lassen und erst im Anschluß mittels Konvertiererroutinen das SGML-Format zu erzeugen. Dieser Weg ist allerdings immer durch die geringere Komplexität der Textverarbeitungssysteme gegenüber Satzrechnern beschränkt und setzt voraus, daß der Verlag gerade über *die* Konvertiererroutine verfügt, die zu dem Textverarbeitungssystem eines Autors paßt. Da der Autor bestimmte Vorschriften beim Umgang mit seinem Textverarbeitungssystem berücksichtigen muß, wie oben (Beispiel: KONSA)

gezeigt, muß der Verlag x Vorschriften und x Konvertierroutinen zur Verfügung haben, will er nicht Autoren mit dem "falschen" Textverarbeitungssystem ausschließen. In einem strengeren Sinne kann man hier nur von einem SGML-System sprechen, wenn der Verlag über einen Parser verfügt, der das konvertierte Dokument "durchcheckt".¹⁰⁹

In allen Fällen, in denen der Autor nicht über eine Software mit implementierter Dokumenttypdefinition verfügt, und das wird die nächste Zeit ganz gewiß die Regel sein, teilt der Verlag dem Autor über Autorenrichtlinien mit, wie er ein Manuskript angefertigt haben möchte. Autorenrichtlinien sind entweder synonym mit Auszeichnungsrichtlinien oder bilden eine Teilmenge davon. Auch hier hängt die Frage der Zumutbarkeit mit der der Akzeptanz zusammen. Es ist in der Diskussion, wie anspruchsvoll bzw. aufwendig Autorenrichtlinien für die logische Auszeichnung sein dürfen. Die Variante STRUKTEXT und auch der Springer Verlag (Heidelberg) setzen auf einfache Autorenrichtlinien. Das hält den Auszeichnungsaufwand für den Autor niedrig und bedeutet, daß der Autor auf dem gewissermaßen kleinsten gemeinsamen Nenner der Autorenrichtlinien aller Verlage arbeitet. Er zeichnet die logischen Elemente aus, die "universell" sind. Dadurch wird ihm die Auseinandersetzung mit komplizierten und von Verlag zu Verlag variierenden Auszeichnungsrichtlinien erspart. Bei diesem Verfahren werden die Mühen der Auszeichnung geteilt. Dieser Weg ist vielleicht auch deshalb "akzeptabel", weil er impliziert, daß der Verlag die Auszeichnungstätigkeit nicht einfach auf den Autor abwälzt, von dem er bereits schon dadurch profitiert, daß er ein weiterverarbeitbares maschinenlesbares Manuskript erhält, ohne für geleisteten Erfassungsaufwand zu zahlen.

Ein Nachteil komplizierter Auszeichnungsrichtlinien ist dadurch gegeben, das sie praktisch mit den verlagsintern gewachsenen komplexen Auszeichnungsrichtlinien kompatibel oder ihnen überlegen sein müssen, wenn sie breite Zustimmung und Anwendung finden wollen. Das Beispiel der Auszeichnungsrichtlinien des AAP-Standards für Formeln mag ein Indiz dafür sein, daß (komplexe) Regelungen komplexer Sachverhalte anfälliger gegen Kritik, weniger konsensfähig sind, was letztlich die Akzeptanz der Regelungen schwächt.

¹⁰⁹ Eine weitere Variante der Autor-Verlagsbeziehung verfolgt der Springer Verlag (Heidelberg), der den Autoren eine verlagseigene Software zur Verfügung stellt, die kaum Einarbeitungsaufwand erfordert und Teile der Dokumenttypdefinition einer Zeitschrift kennt, also *die* Teilstruktur enthält, die ein Autor für einen bestimmten Beitrag auszeichnen soll. Zur Akzeptanz kann im Moment nichts gesagt werden.

Besonders akzeptanzfördernd und "strategisch geschickt", um es einmal salopp zu sagen, erachte ich die SGML-Implementation im Rahmen des DFN. Hier wird ein *zusätzlicher* Dienst im Rahmen der Fachkommunikation *und* technisch-wissenschaftlichen Koproduktion (Autoren, Designer, Konstrukteure etc.) angeboten, den zu nutzen - nebenbei - voraussetzt, Dokumente logisch auszuzeichnen. Der Aspekt der Lastenverteilung spielt hier keine Rolle.¹¹⁰ Sollte DAPHNE eine Tür zum Verlagsbereich durch den Dokumenttyp STRUKTEXT bekommen, würde das die Nutzungsmöglichkeiten und damit die Akzeptanz der "SGML" noch vergrößern.

Weiteren Akzeptanzzuwachs erführe die SGML von den heutigen DCF-Nutzern, wenn sich die IBM entschließen würde, die DCFGML zu einer echten SGML-Implementation mit erforderlichen Verbesserungen der Benutzeroberfläche auszubauen.

5.2 ÜBERGREIFENDE ERGEBNISSE UND OFFENE FRAGEN

In dieser Arbeit wurden Materialien zusammengetragen, die sich wesentlich darum drehten, die soziale Autor - Verlagsbeziehung oder technisch die Beziehung Elektronisches Manuskript - Datenbasis im neutralen Format aufzuklären. Die Beschäftigung mit der SGML stellte den Bezug zu Liefer- und Nutzungsformen her. In einer Graphik habe ich versucht, die Komplexität eines elektronischen Publiziersystems darzustellen, was vielleicht besser geeignet ist, die Ergebnisse dieser Arbeit zusammenzufassen als viele Worte. Es wurde versucht, die Komplexität der Textelemente ebenso zu berücksichtigen (von der Ersterfassung bis zum Angebot), wie die technische Ausstattung, die an den verschiedenen Stellen im Publiziersystem vorhanden sein muß. Auch die Software wurde mit eingebracht unter besonderer Berücksichtigung SGML-spezifischer Programme.

Die Arbeit hat neben der Frage nach der Akzeptanz weitere Fragen zum elektronisch publizierenden Autor aufgeworfen, denen eine anwendungsorientierte Informationswissenschaft nachgehen sollte:

1. Die Arbeit hat deutlich gemacht, daß für empirische Analysen das Konzept des Autors differenziert werden muß. Die Trennung

¹¹⁰ Wie aufwendig das Erstellen des SGML-Dokuments ist, berührt das Forschungsnetz nur von der Seite der Dokumenttypdefinitionen - nicht von der Seite der Softwareunterstützung bei der Erfassung.