

Universität des Saarlandes

Fachrichtung 5.6 - Informationswissenschaft

Peter Weiland

Weiherrstr. 75

66773 Schwalbach

**Optimierung der informationellen Infrastruktur
für die wissenschaftliche Kommunikation**

Abschlussarbeit zur Erlangung des
akademischen Grades eines Magister Artium (M.A.)

Erstgutachter: PD Dr. Ilse Harms

Zweitgutachter: Prof. Dr. Harald H. Zimmermann

Saarbrücken, November 2000

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Saarbrücken, den 15.11.2000

Für meine Eltern. Und Eva.

Inhaltsverzeichnis

1	Einführung	1
1.1	Motivation der Arbeit	1
1.2	Aufbau der Arbeit	2
2	Traditionelle wissenschaftliche Kommunikation	3
2.1	Das Garvey/Griffith Modell der wissenschaftlichen Kommunikation	3
2.2	Formelle und informelle Kommunikation	4
2.3	Beteiligte im Prozess der wissenschaftlichen Kommunikation	5
2.3.1	Wissenschaftler als Autoren und Rezipienten	6
2.3.2	Wissenschaftler als Gutachter	6
2.3.3	Verlage und Fachzeitschriften	6
2.3.3.1	Editorial	7
2.3.3.2	Production	7
2.3.3.3	Marketing	7
2.3.3.4	Fulfillment	7
2.3.3.5	Subsidiary Rights	7
2.3.3.6	Financial Oversight	8
2.3.4	Fachinformationsdienste	8
2.3.5	Bibliotheken	9
2.4	Peer Review	9
2.5	Krise der wissenschaftlichen Literaturversorgung	11
2.5.1	Wachstumsraten der wissenschaftlichen Literatur	11
2.5.2	Die Preispolitik der Verlage	12
2.5.3	Die Geschwindigkeit von gedruckten Publikationen	13
3	Formelle wissenschaftliche Kommunikation über das Internet	14
3.1	Überblick formelle elektronische Kommunikation	14

3.1.1	Ausprägungen des elektronischen Publizierens in der wissenschaftlichen Fachkommunikation	14
3.1.1.1	e-journal	15
3.1.1.2	hybrid paper-electronic journal	15
3.1.1.3	hybrid electronic-paper journal	15
3.1.1.4	electronic working articles	16
3.1.2	Konkrete Implementierungen	16
3.1.2.1	Aktivitäten der Fachverlage	16
3.1.2.2	Preprint-Server und „graue“Literatur	17
3.1.2.3	Forschungsprojekte in Deutschland	19
3.1.3	Kosten für elektronische Publikationen	21
3.2	Vorteile der elektronischen Verfügbarkeit von Dokumenten	22
3.2.1	Verfügbarkeit	22
3.2.2	Schnellere Publikation	22
3.2.3	Platzbedarf	22
3.2.4	Recherchemöglichkeiten	23
3.2.5	Aktualisierung von elektronischen Publikationen	23
3.2.6	Förderung der informellen Kommunikation	23
3.2.7	Multimedia und personalisierte Dienste	24
3.3	Problemfelder der elektronischen wissenschaftlichen Kommunikation	24
3.3.1	Information Retrieval im Internet	25
3.3.2	Versionskontrolle bei elektronischen Dokumenten	25
3.3.3	Zitation elektronischer Dokumente	26
3.3.4	Zugang zu elektronischen Aufsätzen	26
3.3.5	Archivierung und langfristige Speicherung elektronischer Publikationen	27
3.3.6	Interoperabilität	28
3.4	Ein Szenario für die formelle wissenschaftliche Kommunikation über das Internet	31
4	Infrastruktur für die wissenschaftliche Kommunikation über das Internet	33
4.1	Objekte und Dokumente	33
4.2	Metadaten	35
4.2.1	Funktion von Metadaten	35
4.2.2	Dublin Core als ein Beispiel für deskriptive Metadaten	36
4.2.2.1	Bisherige Entwicklung	36

4.2.2.2	Unqualifiziertes vs. qualifiziertes Dublin Core . . .	38
4.2.2.3	Speicherung von Dublin Core Metadaten	41
4.2.2.4	Das Nordic Metadata Project als Beispiel für eine Implementierung	42
4.2.3	Interoperabilität auf Metadaten-Ebene	43
4.2.3.1	Schemas	43
4.2.3.2	Registries und Crosswalks	44
4.2.3.3	Modulare Metadaten: Das Warwick Framework . .	46
4.2.4	Resource Description Framework	46
4.2.4.1	Das Datenmodell des Resource Description Frame- work	47
4.2.4.2	Die Syntax des Resource Description Framework .	48
4.2.4.3	RDF Schema	50
4.2.4.4	Werkzeuge für den Einsatz von RDF	50
4.2.5	Technische Infrastruktur	50
4.2.5.1	Suchmaschinen	51
4.2.5.2	Metadaten-Editoren	52
4.2.6	Diskussion	52
4.3	Identifikation und Lokation von Web-Ressourcen	53
4.3.1	Reichweite von Identifiers	53
4.3.2	Warum die URL kein Identifier ist	54
4.3.3	Anforderungen an Identifiers	54
4.3.4	Uniform Resource Names	55
4.3.4.1	Definition	55
4.3.4.2	Framework für die Anwendung von URNs	56
4.3.4.3	Syntax der URN	56
4.3.4.4	URN-Resolution	57
4.3.4.5	Der URN-Generator des Nordic Metadata Projects .	58
4.3.5	Digital Object Identifier (DOI)	59
4.3.5.1	Ursprung und Entwicklung	59
4.3.5.2	Was identifiziert ein DOI?	60
4.3.5.3	Das DOI-System und seine Komponenten	61
4.3.5.4	Enumeration	62
4.3.5.5	Description	62
4.3.5.6	Resolution	63

4.3.5.7	Policies	65
4.3.5.8	Registrierung von DOIs	67
4.3.5.9	Verbreitung des DOI-Systems	68
4.3.6	Diskussion	68
4.4	Reference Linking und Citation Linking	70
4.4.1	Grundannahmen für den Einsatz von Reference Linking	70
4.4.2	Auflösung einer Referenz zu einem Identifier	72
4.4.2.1	Statische Verlinkung	72
4.4.2.2	Dynamische Verlinkung	74
4.4.2.3	Adaption an die Umgebung	75
4.4.3	Citation Indexes und Citation Linking	76
4.4.4	Implementierungen und Werkzeuge	77
4.4.4.1	Web of Science	78
4.4.4.2	Reference Linking mit DOIs (CrossRef)	80
4.4.4.3	SLinkS	82
4.4.4.4	Research Index	85
4.4.4.5	SFX	86
4.4.4.6	Distributed Link Service	87
4.4.4.7	Andere Ansätze	88
4.4.5	Diskussion	88
4.5	Management des Zugangs zu digitalen Objekten	90
4.5.1	Allgemeines Modell der Autorisierung	90
4.5.2	Technologien für das Zugangsmanagement	91
4.5.2.1	IP-Filter	91
4.5.2.2	Benutzername/Passwort	92
4.5.2.3	Credential-Based	92
4.5.2.4	Proxy-Server	93
4.5.3	Diskussion der vorgestellten Ansätze	93
4.6	Weitere Aspekte	94
4.6.1	Usability	94
4.6.2	Retro-Digitalisierung	94
5	Andere Ansätze für die wissenschaftliche Kommunikation über das Internet	96
5.1	Modulare Struktur für elektronische wissenschaftliche Artikel	96
5.2	Die Modelle von Hurd	97

5.2.1	The Unvetted Model	97
5.2.2	The Collaboratory Model	98
5.3	Scholarly Skywriting	99
6	Zusammenfassung und Fazit	100
6.1	Zusammenfassung	100
6.2	Fazit	102

Tabellenverzeichnis

4.1	Das Dublin Core Element Set, Version 1.1 (REC-DCES, 1999)	39
4.2	Mapping zwischen CORC und Dublin Core (Ausschnitt), MetaForm (1997–2000)	45
4.3	Syntax der URN	56
4.4	Fiktives Beispiel der Kodierung einer ISBN in eine URN	57
4.5	Syntax des DOI nach ANSI/NISO Z39.84-2000 (Paskin, 2000, Ap- pendix 1.2)	62
4.6	Beispiel für die URL einer Fachzeitschrift unter http://www.publisher.com , basierend auf dem Volume und der Startseite des Artikels (Hellman, 1999b, S. 70)	82
4.7	Taxonomy of features for designing digital libraries (Kengeri u. a., 1999)	95

Abbildungsverzeichnis

2.1	Garvey/Griffith-Modell der wissenschaftlichen Kommunikation (Hurd, 1996, S. 11)	4
3.1	Publikationsprozess mit einem Nebeneinander von elektronischen Zeitschriften und Dokumentenservern	31
4.1	Teile eines digitalen Objekts (Arms, 1995)	34
4.2	Meta-object (Arms, 1995)	34
4.3	Organisationale Struktur der Dublin Core Metadata Initiative (Weibel, 2000b)	37
4.4	Unkontrollierte Qualifikation vs. Interoperabilität (Lagoze, 2000) . . .	41
4.5	Abbildung von zwei Vokabularien aufeinander über eine <i>Semantic Layer</i> (Heery u. a., 2000)	46
4.6	Graphendarstellung eines einfachen Tripels	48
4.7	Graphendarstellung eines Tripels mit komplexem Wert für ein Attribut	48
4.8	Aussage über eine Aussage	48
4.9	Globale Auflösung von URNs mithilfe des Resolver Discovery System	58
4.10	Lokale Auflösung von URNs mithilfe eines Proxy-Servers	59
4.11	Beziehungen zwischen den verschiedenen Typen geistigen Eigentums (Rust und Bide, S. 23)	61
4.12	Auflösung eines DOI aus Sicht des Nutzers (Bide, 1997)	64
4.13	<i>Multiple resolution</i> eines DOI (Paskin, 2000, Seite 6-2)	65
4.14	Architektur des Handle-Systems (Sun und Lannom, 2000)	66
4.15	Allgemeines Modell des Reference Linking (Caplan und Arms, 1999)	71
4.16	Kodierung von bibliographischen Informationen mit Dublin Core in XML (Apps und MacIntyre, 2000)	74
4.17	Citation Linking	77
4.18	Sammlung von Metadaten und DOI/URL-Registrierung	81

4.19	Abfrage der DOI-X Datenbank	81
4.20	Zentralisierte Linking-Architektur	83
4.21	Private Linking-Architektur	84
4.22	Intranet Linking-Architektur	85
4.23	Modell des Zugangs-Managements (vereinfacht nach Arms, 1999a, S.34)	91
5.1	The Unvetted Model (Hurd, 1996, S. 28)	98
5.2	The Collaboratory Model (Hurd, 1996, S. 30)	98

Kapitel 1

Einführung

1.1 Motivation der Arbeit

Seit die *Philosophical Transactions of the Royal Society of London* 1665 zum ersten Mal erschienen sind, stellen wissenschaftliche Fachzeitschriften das wichtigste Medium der formellen wissenschaftlichen Kommunikation dar. Bedingt durch die Möglichkeiten des Internet und den Problemen des traditionellen Publikationssystems, steigt das Angebot und die Nachfrage für online-verfügbare wissenschaftliche Fachpublikationen.

Die Arbeit beschreibt, wie die wissenschaftliche Kommunikation durch den Einsatz elektronischer Publikationen verbessert werden kann und welche Anforderungen sich daraus an die zugrundeliegende Infrastruktur ergeben. Ausgehend von der wissenschaftlichen Fachzeitschrift werden dabei diejenigen Problemfelder besprochen, die eine Transformation von gedruckten Publikationen zu elektronisch verfügbaren Publikationen mit sich bringen. Ausgehend von den Problembereichen und den Mehrwertfunktionen von elektronischen Publikationen wird ein Szenario für eine optimierte wissenschaftliche Kommunikation über das Internet entworfen. Dem Szenario werden konkrete Lösungen aus Forschung und Praxis gegenübergestellt, und es wird untersucht, inwiefern die einzelnen Komponenten des Szenarios mit verfügbaren Technologien realisierbar sind.

Der Schwerpunkt der Arbeit liegt auf der technischen Infrastruktur, die für eine effektive wissenschaftliche Kommunikation über das Internet notwendig ist. Die ökonomischen, rechtlichen und sozialen Implikationen eines Übergangs von gedruckten zu elektronischen Publikationen werden dabei nur gestreift.

1.2 Aufbau der Arbeit

Kapitel 2 beschreibt die traditionelle papierbasierte wissenschaftliche Kommunikation sowie die beteiligten Akteure und ihre Rollen innerhalb dieses Prozess. Darüber hinaus wird auf die Qualitätskontrolle in der Wissenschaft und die sich seit einigen Jahren abzeichnende Krise in der wissenschaftlichen Literaturversorgung eingegangen.

Kapitel 3 gibt einen Überblick über existierende Ausprägungen der wissenschaftlichen elektronischen Kommunikation und beschreibt die Mehrwertfunktionen und die zu lösenden Probleme. Darauf aufbauend wird ein Szenario für eine optimierte wissenschaftliche Kommunikation über das Internet entwickelt.

Kapitel 4 stellt dem Szenario konkrete (existierende) Technologien gegenüber. Es wird beschrieben, wie Objekte in digitalen Kollektionen aufgebaut sind, die Möglichkeiten der Auszeichnung von digitalen Objekten mit Metadaten und Technologien zur Identifikation und Lokalisierung von Web-Ressourcen. Darüber hinaus wird das Citation/Reference Linking als eine wesentliche Mehrwertfunktion der elektronischen Kommunikation detailliert erläutert. Möglichkeiten zur Regelung des Zugangs zu elektronischen Kollektionen schliessen dieses Kapitel ab.

Kapitel 5 stellt einige Konzepte für die wissenschaftliche Kommunikation über das Internet vor, die insbesondere den informellen Austausch von Wissenschaftlern fördern sollen.

Kapitel 6 gibt eine Zusammenfassung und Bewertung der vorgestellten Infrastruktur.

Kapitel 2

Traditionelle wissenschaftliche Kommunikation

2.1 Das Garvey/Griffith Modell der wissenschaftlichen Kommunikation

Garvey und Griffith (1979) beschreiben den Prozess der wissenschaftlichen Kommunikation ausgehend von empirischen Untersuchungen bei Psychologen und führen dabei die verschiedenen Phasen von der Forschungsidee bis zur Assimilation der Arbeit in das allgemein akzeptierte Wissensgut auf. Abbildung 2.1 stellt eine Generalisierung des beschriebenen Prozesses dar.

Der Ablauf eines Kommunikationsprozesses in der Wissenschaft stellt sich nach diesem Modell folgendermaßen dar: Während eines Forschungsprojektes kommuniziert ein Wissenschaftler mit Fachkollegen und veröffentlicht sogenannte *preliminary reports* (z.B. Forschungsberichte eines Fachbereichs). Die vollendeten Forschungsergebnisse werden dann in Seminaren und Kolloquien vorgestellt oder auch - in einem größeren Rahmen - auf Konferenzen, wobei der ausgearbeitete Vortrag dann in dem Proceedingsband der jeweiligen Konferenz erscheint. Der nächste Schritt ist die Einreichung eines Manuskripts bei einem Verlag (oder auch einer Fachgesellschaft) zur Begutachtung; parallel hierzu verteilt der Wissenschaftler Preprints seiner Forschungsarbeit an Kollegen. Ist das Manuskript für eine Veröffentlichung in einer Fachzeitschrift akzeptiert worden, dann erscheint es in der Liste der akzeptierten Manuskripte und später dann als Artikel in dieser Fachzeitschrift. Kurz nach der Veröffentlichung erscheint der Titel dann in den *Current Contents*, bei Abstracting-Diensten, in den *Annual Reviews* und schließlich in den Literaturangaben von anderen Artikeln.

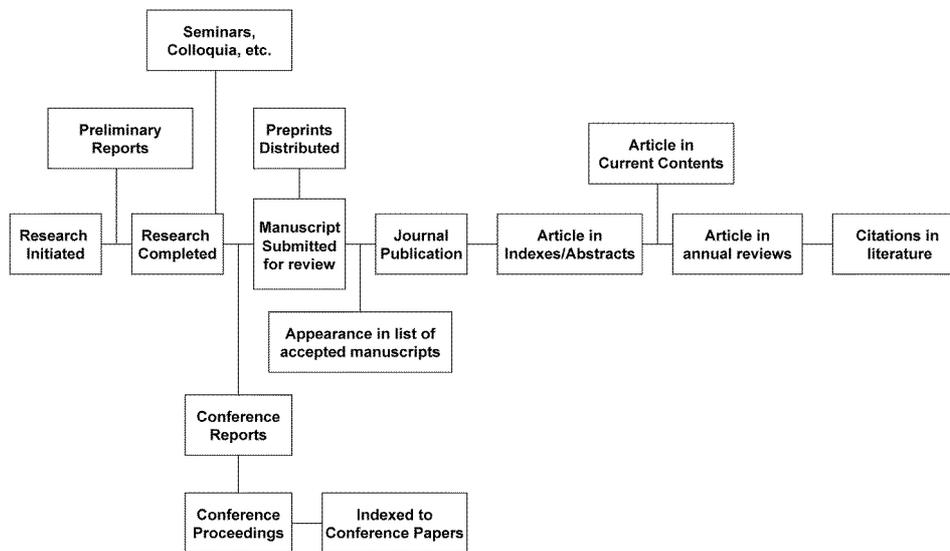


Abbildung 2.1: Garvey/Griffith-Modell der wissenschaftlichen Kommunikation (Hurd, 1996, S. 11)

2.2 Formelle und informelle Kommunikation

Der Fokus des Garvey/Griffith-Modells liegt auf der *formellen* Kommunikation; die *informelle* Kommunikation während der konkreten Forschungsphase spielt hier eine untergeordnete Rolle. Nach Felt u. a. (1995, S. 67) umfasst die informelle Kommunikation den privaten Meinungs-austausch zwischen Wissenschaftlern, sei es persönlich, fernmündlich oder per Email. Die formelle Kommunikation hingegen geht immer mit einer Veröffentlichung einher, da hier die Nachprüfbarkeit von Ergebnissen im Vordergrund steht. Ausprägungen der formellen wissenschaftlichen Kommunikation sind z.B. Monographien, Zeitschriftenartikel und Forschungsberichte.

Bei der formellen Kommunikation gibt es hinsichtlich des Stellenwerts von Veröffentlichungen teilweise gravierende Unterschiede zwischen den Disziplinen: Kling und Covi (1997) untersuchten die Kommunikationsgewohnheiten von Wissenschaftlern und Studenten in den Disziplinen Informatik, Molekularbiologie, Literaturwissenschaft und Soziologie:

- Für Informatiker machen Artikel in Fachzeitschriften und Proceedingsbänden den Hauptbestandteil der formellen Kommunikation aus. Diese Veröffentlichungen erscheinen in den Publikationslisten der jeweiligen Wissenschaftler. Dane-

ben werden noch sogenannte Technical Reports publiziert, die von der jeweiligen Fakultät herausgegeben werden. Diese sind in der Regel keinem Peer Review (siehe 2.4) unterzogen worden und haben auch keine karrierefördernde Wirkung; sie dienen vor allem der Rechtfertigung von Drittmitteln.

- In der Molekularbiologie stellen Artikel in wissenschaftlichen Fachzeitschriften die wichtigste Form der Verbreitung von Forschungsergebnissen dar. Konferenzen in dieser Disziplin dienen eher der Förderung des informellen Austausches, da wichtige Resultate für die Publikation in den einschlägigen Fachzeitschriften vorgehalten werden.
- Im Gegensatz zu den vorgenannten Disziplinen stellen die Literaturwissenschaften einen Sonderfall dar. Selten bauen Artikel in Zeitschriften auf anderen Artikeln auf, die tatsächliche Kommunikation läuft eher über einen informellen Austausch mit Kollegen, die das gleiche Gebiet bearbeiten. Der Arbeitsablauf wird in Form von Projekten beschrieben, wobei die Gründe für das Starten eines Projektes unterschiedlich sein können: Vorträge, die Entdeckung von neuen Quellen oder auch die Zusammenarbeit von verschiedenen Literaturwissenschaftlern, die ein bestimmtes Werk kommentieren. Die Arbeit an Büchern kann mehrere Jahre in Anspruch nehmen. Im Allgemeinen sind die Literaturwissenschaften sehr viel weniger von einer schnellen Publikation abhängig. Bücher spielen für die Reputation eines Wissenschaftlers eine sehr viel größere Rolle als Veröffentlichungen in Zeitschriften.
- In der Soziologie wird die formelle Kommunikation über peer-reviewed Artikel und Bücher abgewickelt. Beiträge zu Konferenzen spielen in der formellen Kommunikation kaum eine Rolle und werden auch nur selten zitiert. Zur Festigung der Reputation eines Wissenschaftlers ist das Verfassen eines Buches notwendig, das keine überarbeitete Fassung seiner Dissertation darstellt.

2.3 Beteiligte im Prozess der wissenschaftlichen Kommunikation

An dem im Garvey/Griffith-Modell dargestellten Prozess sind verschiedene Instanzen beteiligt, deren Funktionen im folgenden näher untersucht werden.

2.3.1 Wissenschaftler als Autoren und Rezipienten

Wissenschaftler nehmen in diesem Prozess eine Sonderrolle ein, da sie Autor und Rezipient zugleich sind. Böhle (1997, S. 398) beschreibt dies als Kette: Arbeitsprozess - Informationsproduktion - Vermittlungsinstanzen - Informationsaneignung - Arbeitsprozess. Die Autoren und Rezipienten stellen in dieser Kette die Endpunkte dar.

Der Publikationsprozess wandelt privates in öffentliches Wissen um; erst mit einer Veröffentlichung kann ein Wissenschaftler Forschungsergebnisse für sich beanspruchen. Merton (1988, S. 606) beschreibt diese Eigenheit der wissenschaftlichen Kommunikation als „seeming paradox“:

I propose the seeming paradox that in science, private property is established by having its substance freely given to others who might want to make use of it.

Lederberg (1993) verweist auf die hohen Qualitätsansprüche an wissenschaftliche Publikationen. Wissenschaftler verbürgen sich mit ihrem Namen für die Richtigkeit der getroffenen Aussagen und können von diesen nicht einfach zurücktreten.

Auf der anderen Seite ist das Publizieren von Forschungsergebnissen unabdingbar („publish or perish“)¹, da nur die Anerkennung durch die Fachöffentlichkeit die Karriere eines Wissenschaftlers langfristig sichert. Anerkennung wird hier ausgedrückt durch Veröffentlichungen in wichtigen Fachzeitschriften und Zitierung in den Arbeiten anderer Wissenschaftler. Die Vergabe von Drittmitteln und Stellen wird oft von der Anzahl und der Qualität der Veröffentlichungen bestimmt, wobei produktivere Wissenschaftler im Allgemeinen mehr Anerkennung erreichen als weniger produktive, unabhängig von der Qualität der Arbeiten (Merton, 1968).

2.3.2 Wissenschaftler als Gutachter

Neben der oben dargestellten Autor/Nutzer-Rolle nehmen etablierte Wissenschaftler auch oft eine Gutachter-Rolle für Veröffentlichungen anderer Wissenschaftler ein. Der Gutachter-Prozess, das sogenannte *Peer Review*, wird in 2.4 ausführlich erläutert.

2.3.3 Verlage und Fachzeitschriften

Fisher (1996) nennt als Aufgabengebiete der Verlage *Editorial*, *Production*, *Marketing*, *Fulfillment*, *Subsidiary Rights* und *Financial Oversight*.

¹In einem 1996 erschienenen Artikel im „The Scientist“ versucht Garfield, dem Ursprung dieses Terminus auf die Spur zu kommen - allerdings ohne Erfolg.

2.3.3.1 Editorial

Verlage treffen Vereinbarungen mit Herausgebern von Fachzeitschriften, die für die Qualitätskontrolle einer wissenschaftlichen Fachzeitschrift verantwortlich sind. Diese Herausgeber werden normalerweise in Form von Personal und Sachmitteln von einem Verlag unterstützt.

2.3.3.2 Production

Die zweite Aufgabe des Verlages ist die konkrete „Herstellung“ des Journals, i.e. Design, Layout, Korrektur, Satz, Druck und Binden. Darüber hinaus werden Termine und Kosten festgelegt und überwacht.

2.3.3.3 Marketing

Verlage möchten ihre Zeitschriften „bekannt“ machen. Dies kann auf vielfältige Weise geschehen, z.B. durch Prospekte und Kataloge, die an potenzielle Abonnenten (Individuen oder Institutionen) gesendet werden, Präsenz auf Messen oder Konferenzen, Pressemitteilungen, Anzeigenschaltung etc..

2.3.3.4 Fulfillment

Verlage bearbeiten alle Bestellungen für eine Fachzeitschrift, seien es Abonnements, Einzelbestellungen oder auch Anforderungen für zurückliegende Ausgaben. Zu den weiteren Aufgaben in diesem Bereich gehört auch die Verteilung bzw. Versendung der Exemplare an die Kunden. Darüber hinaus werden auch Abkommen mit Distributoren geschlossen, die die Zeitschriften weiterverkaufen.

2.3.3.5 Subsidiary Rights

Neben dem Direktverkauf der Fachzeitschriften schließt ein Verlag mit anderen Distributoren Abkommen hinsichtlich weiterer Verwertungsrechte wie Nachdrucke, Übersetzungen, Mikrofiche, CD-ROM oder auch Volltext über Datenbanken. Darüber hinaus werden die für die Disziplin relevanten Fachinformationsdienste kontaktiert, damit die Fachzeitschrift in die Produkte der Informationsanbieter integriert wird. Um möglichen Urheberrechtsverletzungen begegnen zu können, muss genau Buch über alle Verträge geführt werden.

2.3.3.6 Financial Oversight

Der Verlag als Unternehmen ist verantwortlich für die finanzielle Seite einer Fachzeitschrift und trägt das gesamte Risiko.

Neben den Aufgaben des Verlages nimmt auch die einzelne Fachzeitschrift wichtige Rollen in der wissenschaftlichen Kommunikation ein. Die Fachzeitschrift mit ihren Artikeln bietet „*a system for formal, public and orderly communication among scientists*“ (Garvey, 1979, S. 69). Formal ist sie in dem Sinne, dass die eingereichten Manuskripte vor der Publikation sorgfältig begutachtet und überarbeitet werden. Darüber hinaus können sie zitiert und eindeutig identifiziert werden. Öffentlich sind Fachzeitschriften, da jedermann Manuskripte einreichen kann und die Fachzeitschriften für jedermann verfügbar sind, sei es über eine Bibliothek oder direkt über eine Subskription. „*Orderly communication*“ ist dadurch sichergestellt, dass Artikel auf Basis des wissenschaftlichen Gehaltes ausgewählt werden: Die Forschungsergebnisse, über die berichtet wird, müssen regulär zustande gekommen sein und die Ergebnisse müssen Relevanz für den wissenschaftlichen Fortschritt haben.

In einer Analyse von Smith (1997) wird die wissenschaftliche Anerkennung als eine wesentliche Aufgabe der wissenschaftlichen Fachzeitschrift gesehen. Daneben sieht Smith auch noch die eher versteckten Rollen, wie die Definition eines Forschungsgegenstandes durch die Herausgeber und ihre inhaltlichen Entscheidungen, die Definition einer *community*, i.e. Gemeinschaft von Wissenschaftlern, die an ähnlichen Themen arbeiten (insbesondere bei gerade entstehenden Forschungszweigen) und auch die Archivierung von wissenschaftlichen Ergebnissen, wobei hier weniger die tatsächliche physische Archivierung gemeint ist - die Aufgabe der Bibliotheken ist - sondern vielmehr die Aufzeichnung des Fortschritts in einem Wissensgebiet.

2.3.4 Fachinformationsdienste

Fachinformationsdienste stellen ein Bindeglied zwischen dem Wissenschaftler als Autor und dem Wissenschaftler als Rezipienten dar, da sie verteilt liegende Informationen zu einem bestimmten Thema zusammenführen und den aktuellen Stand eines Wissensgebiets repräsentieren können. Zu den klassischen Informationsdiensten gehören z.B. die Current Contents, die wöchentlich die Inhaltsverzeichnisse von Fachzeitschriften in verschiedenen Disziplinen wiedergeben. Daneben gibt es die Titellisten, die Beiträge aus Periodika nach den sinntragenden Wörtern im Titel alphabetisch anordnen. Ein bekannter Dienst sind Bibliographien, die nach unterschiedlichsten Gesichtspunk-

ten angeordnet werden können. Eine fortgeschrittenere Form des Zugangs stellen die sogenannten Zitierindexe (engl. citation indexes) dar: Ausgehend von einem bereits bekannten Artikel lassen sich auch alle diejenigen Artikel finden, die auf diesen bekannten Artikel verweisen, d.h. der Nutzer hat also die Möglichkeit, neuere Artikel, die den bekannten Artikel zitieren, zu finden. Zitierindexe wurden 1961 eingeführt vom *Institute for Scientific Information (ISI)* (Ockenfeld, 1997, vgl.). Seit 1997 gibt es mit dem *Web of Science* des ISI eine Hypertext-Variante der Zitierindexe im World Wide Web (siehe 4.4.4.1).

Ein weiteres Hilfsmittel zur Recherche stellen Referatedienste und Register dar. Sie bieten Informationen über Veröffentlichungen, so dass ein Nutzer zum einen entscheiden kann, ob die tatsächliche Beschaffung einer Abhandlung lohnenswert ist, und zum anderen lässt sich durch die sachliche Erschließung herausfinden, ob es zu einem bestimmten Gebiet überhaupt Veröffentlichungen gibt. Die Chemical Abstracts sind ein Beispiel für einen Referatedienst in der Chemie, der sich durch ein komplexes Registersystem auszeichnet (Ockenfeld, 1997, S. 271-273).

Neben den beschriebenen „traditionellen“ Fachinformationsdiensten, die mittlerweile vor allem in elektronischer Form in Gebrauch sind, existieren eine Vielzahl von Referenzdatenbanken, die eine gezielte Suche nach Veröffentlichungen ermöglichen. Zusätzliche Dienste sind die sogenannten Profildienste, die - mit dem Interessenprofil eines Wissenschaftlers versehen - elektronisch über neue Artikel informieren, die auf das eingespeiste Profil passen (Ockenfeld, 1997, S. 274).

2.3.5 Bibliotheken

In der wissenschaftlichen Kommunikation sind die Bibliotheken die Zugangsanbieter zur wissenschaftlichen Fachliteratur. Sie erwerben Publikationen, sei es in Form von Monographien oder Zeitschriftenabonnements, erschließen die erworbenen Publikationen nach bestimmten Kriterien, handhaben den Zugang zu den Publikationen und konservieren die gesammelten Publikationen für die Nachwelt. Darüber hinaus sind sie die Hauptabnehmer wissenschaftlicher Literatur, was sie zu einer festen Größe in der Absatzplanung der Verlage macht.

2.4 Peer Review

Garfield (1986) beschreibt das Peer Review, die Qualitätskontrolle für Artikel in wissenschaftlichen Fachzeitschriften, als mehrstufigen Prozess. In einem ersten Schritt bestimmt der Herausgeber, ob ein eingereichter Artikel mit dem Stil und dem Ab-

deckungsgrad der Zeitschrift vereinbar ist. Der zweite Schritt ist die Auswahl geeigneter Gutachter. Hier gibt es unterschiedliche Modelle: Bei einigen Fachzeitschriften werden die Autoren aufgefordert, Gutachter zu benennen, andere Herausgeber wählen die Gutachter aus den im Artikel zitierten Autoren, oder es werden anerkannte Experten eines Fachgebietes als Gutachter benannt. Teilweise werden auch bibliographische Datenbanken für die Suche nach Gutachtern herangezogen. Ebenso wie der Gutachterausswahl-Prozess variiert auch die Anzahl der Gutachter von Zeitschrift zu Zeitschrift, wobei zwei den Normalfall darstellt. Nachdem eine Auswahl getroffen wurde, wird das Manuskript zusammen mit Instruktionen und Formblättern für Kommentare und Empfehlungen den Gutachtern zugestellt, und nach wenigen Wochen antworten die Gutachter und empfehlen entweder die Ablehnung oder die Publikation des Manuskripts oder sie machen Vorschläge für Modifikationen. Drei Kriterien sind für die Bewertung eines Manuskripts ausschlaggebend. Es sollte *sound* sein, d.h., dass die eingesetzten Forschungsmethoden zuverlässig sind, gültige Schlussfolgerungen getroffen wurden und keine logischen Fehler enthalten sind. Zum zweiten sollte es *original* sein, d.h. die Ergebnisse werden zum ersten Mal veröffentlicht. Das dritte Kriterium ist, ob das Manuskript *significant* ist. Dies bedeutet, dass eine eingereichte Arbeit neue Perspektiven oder potenziell wichtige Beobachtungen enthält. Die endgültige Entscheidung, ob ein Manuskript tatsächlich veröffentlicht wird, trifft dann der Herausgeber der Zeitschrift auf Basis der (manchmal auch widersprüchlichen) Urteile der Gutachter.

Dieses System der wissenschaftlichen Qualitätskontrolle führt zu einer Rangfolge der Fachzeitschriften: Diejenigen mit den höchsten Ablehnungsraten für eingereichte Artikel rangieren am höchsten, da sie nur die besten Arbeiten in einem Gebiet veröffentlichen. Allerdings gibt es große Unterschiede zwischen den einzelnen Disziplinen: Prestigereiche Zeitschriften in den Geistes- oder Sozialwissenschaften lehnen teilweise 80-90% der eingereichten Manuskripte ab, in anderen Disziplinen, wie der Mathematik und Physik, sind die Prozentzahlen sehr viel kleiner (Harnad, 1996, S. 110).

Viele Wissenschaftler schätzen das Peer Review als eine wichtige Instanz im Prozess der wissenschaftlichen Kommunikation, dennoch gibt es einige Kritikpunkte an diesem System:

- Peer Review ist langsam und fehlerbehaftet: Die Gutachter haben häufig eine hohe Arbeitsbelastung und sind nicht in der Lage, die Gutachten zeitig und in der gebotenen Qualität zu liefern (Harnad, 1996, S. 112).

- Die Gutachter sind voreingenommen: Autoren beklagen sich, dass Manuskripte von weniger renommierten Autoren und Instituten eher abgelehnt werden als solche von renommierten (van Rooyen, 1998).
- Die Gutachter sind gegenüber neuen, innovativen Ideen oft nicht aufgeschlossen und lehnen diese eher ab als konservative Ansätze. Garfield (1977) nennt dies die „Newcomb variety“; Simon Newcombe hat wenige Monate vor dem ersten Flug der Gebrüder Wright bewiesen (und darüber veröffentlicht), dass eine Flugmaschine unmöglich ist.

Die lange Dauer des Peer Review ist oftmals mitverantwortlich für den langen Zeitraum von der Einreichung bis zur Veröffentlichung eines Manuskripts. Wood (1998) sieht das *Online Peer Review* als eine mögliche Lösung zur Kosten- und Zeitersparnis an, da die Manuskripte inklusive aller Graphiken per Email an die Gutachter geschickt werden und auch die Bewertung wieder per Email an den Verlag gesendet wird. Somit kann auf die zeitraubende Vervielfältigung und die Versendung per Post verzichtet werden. Die Probleme beim Online Peer Review liegen in den unterschiedlichen Dateiformaten, die von Autoren benutzt werden, Sicherheitsbedenken und generellen Problemen bei der Handhabung von Email.

Eine andere Alternative ist ein sogenanntes offenes online Peer Review, bei dem praktisch jeder (namentlich) ein im Web veröffentlichtes Manuskript bewerten kann (Butler, 1999).

2.5 Krise der wissenschaftlichen Literaturversorgung

2.5.1 Wachstumsraten der wissenschaftlichen Literatur

Seit einiger Zeit spricht man von einer Krise in der wissenschaftlichen Literaturversorgung, die sich dahingehend auswirkt, dass den Bibliotheken die finanziellen Mittel zur Beschaffung der relevanten Publikationen (insbesondere Fachzeitschriften) fehlen. Ein wichtiger Grund ist die steigende Zahl an Veröffentlichungen. Odlyzko (1996) gibt einige Zahlen an, die die Größenordnung des Wachstums verdeutlichen sollen:

- Die Zahl der pro Jahr veröffentlichten wissenschaftlichen Artikel hat sich in den letzten beiden Jahrzehnten alle 10-15 Jahre verdoppelt. In den Geowissenschaften hat sich nach dem zweiten Weltkrieg die Zahl der Veröffentlichungen alle 8 Jahre über einen ausgedehnten Zeitraum verdoppelt.

- Die Zahl der Abstracts, die in den *Chemical Abstracts* veröffentlicht wurden hat sich von 1945-1980 jedes Jahrzehnt verdoppelt, wobei das Wachstum mittlerweile wieder abgeflacht ist.
- Für die Mathematik ergibt aus diesen Wachstumsraten, dass die Hälfte der gesamten verfügbaren mathematischen Literatur in den letzten 10 Jahren entstanden ist.

2.5.2 Die Preispolitik der Verlage

Neben dem Wachstum in der Anzahl der Veröffentlichungen spielt auch die Preispolitik der wissenschaftlichen Verlage eine große Rolle für die Probleme in der Literaturversorgung. Das Budget der deutschen Universitätsbibliotheken für Erwerbungen beträgt jährlich ungefähr 225 Millionen DM. Im Zeitraum von 1991 bis 1997 wuchs dieses Budget um nominal 1,3% an, wobei sich die Fachzeitschriften in den Geistes- und Sozialwissenschaften um durchschnittlich 27% und in den Natur- und Ingenieurwissenschaften um 77% verteuert haben (Sietmann, 1999). In einem offenen Brief² an sieben Fachverlage im Januar 1999 beklagt die Kommission des Deutschen Bibliotheksinstituts für Erwerbung und Bestandsentwicklung die durchschnittlichen Preiserhöhungen von 1998 auf 1999, die sich zwischen 13 und 27% bewegen.

Sietmann (1999) und Odlyzko (1996) weisen auf die Besonderheiten des Marktes für wissenschaftliche Fachkommunikation hin, der eigentlich keinen Markt darstellt. Die Nutzer von Fachpublikationen müssen für diese nicht bezahlen, da diese in der Regel von Bibliotheken angeschafft werden. Daher gibt es für sie keinen Grund, auf die Veröffentlichung in hochpreisigen (und unter Umständen prestigereichen) Zeitschriften zu verzichten und somit Druck auf die Verlage auszuüben. Desweiteren stehen die Verlage in keiner Konkurrenz um Autoren oder Gutachter, da ihre Produkte sehr kleine, spezialisierte Bereiche abdecken. Morton (1997) sieht die Fusionen im wissenschaftlichen Verlagsbereich als eine weitere Ursache für die fehlende Wettbewerbssituation.

Für die Bibliotheken mit ihren stagnierenden Budgets bedeutet die Preispolitik der Verlage zum einen, dass sie Abonnements abbestellen müssen bzw. neue Fachzeitschriften erst gar nicht abonnieren können und zum anderen das Ausweichen auf Dokumentenlieferdienste und Fernleihe. Da die Verlage mit einem fixen Kostenblock operieren, bedeutet die Abbestellung von Abonnements neue Preiserhöhungen für die verbleibenden Kunden (Grötschel und Lügger, 1996).

²http://www.ub.uni-siegen.de/pub/misc/Offener_Brief.pdf, Zugriffsdatum: 25.07.2000

Ein weiterer Aspekt, der zunehmend Widerstand gegen die Verlage auf den Plan ruft, ist die Tatsache, dass die Ergebnisse von öffentlich geförderter Forschung in Form von Fachzeitschriften und Artikeln mit öffentlichen Mitteln von den Verlagen zurückgekauft werden müssen. In den USA versucht die 1997 gegründete *Scholarly Publishing and Academic Resources Coalition (SPARC)*³, eine Allianz von Universitätsbibliotheken, Konkurrenz zu den kommerziellen Verlagen aufzubauen, indem Produkte von Universitätsverlagen und Fachgesellschaften durch garantierte Abnahme durch die beteiligten Bibliotheken unterstützt werden (siehe Rambler, 1999, Butler, 1999).

2.5.3 Die Geschwindigkeit von gedruckten Publikationen

Die Publikation von Artikeln in (gedruckten) wissenschaftlichen Fachzeitschriften war bis vor einiger Zeit die schnellste Möglichkeit zur Verbreitung von Forschungsergebnissen und neuen Ideen. Bedingt durch die große Anzahl von eingereichten Manuskripten, den langwierigen Prozess des Peer Review und Seitenzahlbeschränkungen seitens der Verlage für einzelne Zeitschriften kommt es zu den langen Verzögerungen bis zur tatsächlichen Veröffentlichung eines Artikels in einer Fachzeitschrift (Okerson, 1991). Grötschel und Lügger (1996) sprechen in der Mathematik von 2 bis 3 Jahren bis ein Aufsatz erscheint.

Ginsparg (1994) berichtet, dass in der Hochenergie-Physik bereits seit Mitte der 70er Jahre das primäre Mittel der Kommunikation von Forschungsergebnissen fotokopierte Preprints waren, die per Post an große Verteilerlisten gesendet wurden, zeitgleich mit der Einreichung der Artikel bei den relevanten Fachzeitschriften. Zu dieser Zeit hatten die Fachzeitschriften in dieser Disziplin die Rolle der Filterung von Forschungsergebnissen bereits verloren. Der nächste logische Schritt war das in 3.1.2.2 beschriebene *Los Alamos Physics E-Print Archive*, das zum einen durch die Entwicklung des Computersatzsystems $\text{T}_{\text{E}}\text{X}$, und zum anderen durch die Verbreitung leistungsfähiger Computernetzwerke in den 80er Jahren begünstigt wurde.

³<http://www.arl.org/sparc/>, Zugriffsdatum : 03.08.2000

Kapitel 3

Formelle wissenschaftliche Kommunikation über das Internet

Die oben beschriebenen Probleme der papierbasierten wissenschaftlichen Kommunikation und die Möglichkeiten der Kommunikation über das Internet führen seit Beginn der 90er Jahre zu einem verstärkten Interesse an der Abwicklung der formellen wissenschaftlichen Kommunikation über das Internet.

Im ersten Teil dieses Kapitels wird ein Überblick über die derzeitig vorhandenen Ausprägungen der formellen wissenschaftlichen Kommunikation über das Internet gegeben und konkrete Projekte in diesem Bereich beschrieben. Anschließend wird der Frage nachgegangen, welche konkreten Vorteile sich für den Nutzer durch die elektronische Kommunikation ergeben und welche Probleme zu lösen sind, beziehungsweise welche Anforderungen sich für eine Infrastruktur daraus ergeben. Abschließend wird ein Szenario für die wissenschaftliche Kommunikation entwickelt, das den Rahmen für die in Kapitel 4 beschriebenen Technologien bildet.

Ausgangspunkt der Betrachtungen ist die formelle wissenschaftliche Kommunikation mit der Publikationsform *Artikel*.

3.1 Überblick formelle elektronische Kommunikation

3.1.1 Ausprägungen des elektronischen Publizierens in der wissenschaftlichen Fachkommunikation

Unter einer *elektronischen Publikation* werden im folgenden alle Dokumente gefasst, die auf elektronischem Wege verteilt werden, sei es per Email, über eine elektronische Diskussionsgruppe, als Datei zum Download auf einem FTP-Server oder als Artikel auf einer Web-Seite (vgl. Kling und McKim, 1999). Im Gegensatz zu Böhle (1997) werden Offline-Angebote auf Datenträgern wie CD-ROM oder Diskette ausgeklam-

mert, da sie nicht das Kriterium der elektronischen Distribution erfüllen.

Kling und McKim (1999) weisen auf den unterschiedlichen Gebrauch des Terminus „elektronische Fachzeitschrift“ hin und definieren vier Ausprägungen:

3.1.1.1 e-journal

Ein e-journal besteht aus einer Anzahl von Artikeln und wird auf elektronischem Wege an die Abonnenten verteilt. Es ist - ähnlich wie eine gedruckte Fachzeitschrift - von einem Herausgeber (oder mehreren Herausgebern) bearbeitet worden, und die Artikel wurden einem Peer Review unterzogen.

Bisher gibt es noch verhältnismäßig wenige Fachzeitschriften, die dieser Kategorie zuzuordnen sind. Beispiele hierfür sind die Zeitschrift *Psychology*¹ (siehe 5.3), das *Journal of the Association of Information Systems*² oder auch das *Internet Journal of Science: Biological Chemistry*³.

3.1.1.2 hybrid paper-electronic journal

Die wohl häufigste Ausprägung ist das sogenannte *paper-electronic journal*. Hier ist die Papierform der primäre Distributionskanal, zusätzlich ist die Zeitschrift elektronisch verfügbar.

In 3.1.2.1 werden einige Beispiele dieser Ausprägung vorgestellt.

3.1.1.3 hybrid electronic-paper journal

Im Gegensatz zum vorgenannten *p-e-journal* wird das *electronic-paper-journal* primär auf elektronischen Wege verteilt, die Papierausgabe spielt eine untergeordnete Rolle.

Ein bekanntes Beispiel dieses Typs sind die *Electronic Transactions on Artificial Intelligence (ETAI)*⁴, auf deren Artikel über das Internet zugegriffen werden kann. Zusätzlich zur elektronischen Version gibt es eine auflagenbegrenzte Papierversion, die allerdings nur an eine begrenzte Anzahl von Institutionen, an Wissenschaftler in Ländern mit mangelhafter Internet-Infrastruktur und an die Autoren eines Artikels ausgeliefert wird. Die elektronische Version bietet gegenüber der Papierversion Diskussionsforen, in denen über die einzelnen Beiträge mit anderen Wissenschaftlern diskutiert werden kann.

¹<http://www.cogsci.soton.ac.uk/psychology/>, Zugriffsdatum: 06.11.2000

²<http://jais.aisnet.org/>, Zugriffsdatum: 06.11.2000

³<http://www.netsci-journal.com/>, Zugriffsdatum: 06.11.2000

⁴<http://www.ida.liu.se/ext/etai/>, Zugriffsdatum: 08.08.2000

3.1.1.4 electronic working articles

Neben elektronischen Zeitschriften, die außer der elektronischen Distribution mit papierbasierten Fachzeitschriften korrespondieren (Peer Review, fester Publikationsrhythmus), werden Artikel auch auf privaten oder institutionellen Homepages veröffentlicht, oder sie werden in elektronischen Archiven abgelegt.

3.1.2 Konkrete Implementierungen

Bedingt durch den Boom des Internet seit Mitte der 90er Jahre existiert eine schier unüberschaubare Anzahl an Projekten, die sich mit der elektronischen Distribution von Fachartikeln beschäftigen. Im folgenden kann daher nur ein grober Überblick gegeben werden.

3.1.2.1 Aktivitäten der Fachverlage

Nahezu alle großen wissenschaftlichen Verlagen bieten ihre Produkte (Zeitschriften) mittlerweile in elektronischer Form an. Die folgende Übersicht ist weder vollständig noch repräsentativ, sondern soll nur einen Eindruck vermitteln, welche Dienste angeboten werden.

- Springer bietet unter seinem Service LINK zur Zeit 480 Zeitschriften, 7 Buchserien und 4 Expertensysteme⁵ in elektronischer Form an. Die Zeitschriften sind in 10 Sachgebiete untergliedert, zusätzlich kann man über eine alphabetisch sortierte Aufstellung zugreifen. Daneben kann man die Kollektion mithilfe einer Suchmaschine durchsuchen. Auf die elektronischen Journale kann nur zugegriffen werden, wenn eine Subskription für das Print-Produkt vorliegt. Der Zugang zu LINK wird entweder über den IP-Bereich⁶ gesteuert oder über ein Login und Passwort.
- Elseviers elektronisches Angebot heißt ScienceDirect und bietet den elektronischen Zugang zu allen Fachzeitschriften von Elsevier Science (gegenwärtig mehr als 1100). ScienceDirect ist in drei verschiedene Produkte aufgliedert⁷. Das Einstiegsangebot *ScienceDirect Web editions* bietet Abonnenten ohne

⁵<http://link.springer.de/whatsnew.htm>, Zugriffsdatum 08.08.2000

⁶Von jedem Rechner im Netz der Universität des Saarlandes (IP-Bereich 134.96.XXX.XXX) hat man Zugriff auf die elektronischen Ausgaben der abonnierten Zeitschriften aus dem Springer-Verlag.

⁷http://www.sciencedirect.com/science/page/static/scidir/static_scidir_splash_prodl.html, Zugriffsdatum: 06.11.2000

zusätzliche Kosten den Zugriff auf die elektronischen Ausgaben der abonnierten Fachzeitschriften, wobei der Zugriff auf die letzten 12 Monate beschränkt ist. Die nächste Stufe ist *ScienceDirect OnSite*. Hier werden die ausgewählten Fachzeitschriften lokal beim Abonnenten vorgehalten, d.h. der Zugang zu den elektronischen Zeitschriften kann von der abonnierenden Institution kontrolliert werden. Die Kosten für *ScienceDirect OnSite* setzen sich aus den Kosten für die abonnierten Fachzeitschriften und den Kosten für die Bereitstellung der Software und Wartung zusammen. *ScienceDirect Online* bietet Zugriff auf alle abonnierten Fachzeitschriften über das Internet, wobei neben den aktuellen Ausgaben auch auf die Ausgaben der zurückliegenden 3 bis 4 Jahre (abhängig von der Zeitschrift) zugegriffen werden kann. Der Zugang ist ähnlich wie bei LINK geregelt. Eine Besonderheit ist das Pay-per-view für Artikel aus Zeitschriften, für die kein Abonnement vorliegt.

- John Wiley & Sons bietet mit Wiley Interscience⁸ Zugriff auf über 300 Fachzeitschriften. Jeder Abonnent einer gedruckten Publikation kann auf die jeweilige elektronische Ausgabe ohne zusätzliche Kosten zugreifen. Mit *EarlyView* besteht die Möglichkeit des Zugriffs auf Artikel, die noch nicht in gedruckter Form erschienen sind.

Neben dem Zugriff auf die jeweils aktuelle oder zurückliegende Ausgaben einer Zeitschrift werden von den Verlagen auf elektronischem Wege auch mehr und mehr Artikel veröffentlicht, die erst später in der gedruckten Ausgabe erscheinen (z.B. *Online First* von Springer oder *EarlyView* bei Wiley Interscience). Artikel, die gedruckt noch nicht erschienen sind, werden bei den erwähnten Verlagen mit dem *Digital Object Identifier (DOI)* (siehe 4.3.5) versehen. Dies ist insbesondere dann notwendig, wenn bibliographische Angaben wie Seitenzahlen zum Zeitpunkt des Erscheinens noch nicht verfügbar sind.

3.1.2.2 Preprint-Server und „graue“Literatur

Bedingt durch den langwierigen Publikationsprozess und die Notwendigkeit, Zugriff auf die neuesten Forschungsergebnisse zu haben, wurden schon vor der flächendeckenden Einführung des Internet in Disziplinen wie der Hochenergiephysik sogenannte *Preprints* ausgetauscht (Ginsparg, 1994). Preprints sind im Prinzip nichts anderes als noch nicht begutachtete Beiträge für Fachzeitschriften oder Konferenzen. Die Fort-

⁸<http://www.interscience.wiley.com/>, Zugriffsdatum: 06.11.2000

schritte im Bereich der weltweiten Vernetzung und die Verbilligung der Massenspeicher ermöglichen seit Anfang der 90er Jahre die Installation von dedizierten, elektronischen Preprint-Archiven. Das bekannteste Beispiel ist wohl das *Los Alamos Physics E-Print Archive*⁹, in dem Preprints der Physik, Chemie und Informatik gesammelt und schon vor der möglichen Veröffentlichung in einer Fachzeitschrift zugänglich gemacht werden. Das System arbeitet in allen Bereichen vollkommen automatisiert, jeder Wissenschaftler kann Artikel in den unterstützten Formaten uploaden und bereits aufgespielte Artikel durch neue Versionen ersetzen, wobei sichergestellt ist, dass dies nur durch den ursprünglichen Autor geschehen kann.

In Deutschland gibt es in der Mathematik mit dem *MathNet*¹⁰ (Dalitz u. a., 1997), ein verteiltes Informationssystem für die Mathematik, das mit *MPRESS* ein Preprint-Archiv zur Verfügung stellt. Für den effizienten Zugriff auf die verteilt liegenden Dokumente sind diese über ein an die Mathematik angepasstes Dublin Core Set erschlossen und werden über ein Harvest System erfasst (siehe 4.2.2 und 4.2.5).

Neben den Preprint-Servern gibt es auch Archive zur Sammlung der sogenannten „grauen Literatur“, i.e. Dissertationen, Diplom- und Masterarbeiten, Forschungsberichte etc.. Erwähnt sei hier das Projekt *Dissertationen Online*¹¹, das zum einen Lobbyarbeit bei den Universitäten und Universitätsbibliotheken hinsichtlich der Akzeptanz digitaler Dissertationen leistet, und zum anderen in verschiedenen Teilprojekten Richtlinien und die benötigte Infrastruktur erarbeitet. Von der Kultusministerkonferenz gibt es seit Herbst 1997 einen Beschluss, der die Ablieferung der Dissertation in elektronischer Form zulässt, allerdings verbieten die Promotionsordnungen verschiedener Fakultäten dies noch. Die Vorteile der Ablieferung eines elektronischen Pflichtexemplars liegen auf der Hand: Zum einen ist der Zugriff auf elektronisch verfügbare Arbeiten einfacher und weniger zeitraubend als per Fernleihe oder Dokumentenlieferdienst, und zum anderen entfallen die mitunter hohen Kosten für den Druck, die der Promovende zu tragen hat. In den USA gibt es mit der *National Digital Library of Theses and Dissertations*¹² ein ähnliches Projekt.

Den weltweiten Zugriff auf Forschungsberichte von Informatikfachbereichen bietet die *Networked Computer Science Technical Library (NCSTRL)*¹³. Eine Besonderheit von NCSTRL ist die weltweite, verteilte Struktur, d.h. die Aufsätze werden nicht

⁹<http://www.arxiv.org>, Zugriffsdatum: 07.08.2000

¹⁰<http://www.math-net.de>, Zugriffsdatum 07.08.2000

¹¹http://www.educat.hu-berlin.de/diss_online/, Zugriffsdatum: 10.08.2000

¹²<http://www.ndtld.org>, Zugriffsdatum: 05.08.2000

¹³In Saarbrücken gibt es momentan einen Testzugang unter <http://www.cs.uni-sb.de/ncstrl/>, Zugriffsdatum: 28.08.2000

zentral an einem Ort vorgehalten, sondern auf den lokalen Servern der angeschlossenen Einrichtungen.

3.1.2.3 Forschungsprojekte in Deutschland

Gegenwärtig gibt es international eine fast unüberschaubare Anzahl an Projekten im Bereich des wissenschaftlichen elektronischen Publizierens und des Aufbaus digitaler Bibliotheken. Rusch-Feja (1999) gibt einen Überblick über internationale Projekte im Bereich digitale Bibliotheken. Aus Platzgründen wird hier nur auf Projekte in Deutschland eingegangen.

Neben den beschriebenen Dissertationen Online und MathNet wird in Deutschland im Rahmen des *Global Info* Projekts an einer Verbesserung der Infrastruktur für die wissenschaftliche Kommunikation gearbeitet. Das Bundesministerium für Bildung und Forschung (BMBF) stellt seit 1998 für einen Zeitraum von sechs Jahren insgesamt 60 Millionen DM bereit. Die Ziele der Initiative sind dabei (Rusch-Feja und Becker, 1999):

1. efficient access to world wide information
2. directly from the scientist's desktop
3. while providing the organization for and stimulating fundamental structural changes in the information and communication process of the scientific community.

Ein wesentlicher Bestandteil der Initiative ist die Kooperation zwischen den Gruppen, die am Prozess der wissenschaftlichen Kommunikation beteiligt sind. Dies sind im einzelnen Verlage und Autoren, Distributoren (z.B. Bibliotheken) und die Konsumenten wissenschaftlicher Literatur.

Die Global Info Initiative fördert Projekte in folgenden Schwerpunktbereichen¹⁴:

- Ergänzung und Bearbeitung von Inhalten: Dokumenttypen, Verfahren und Werkzeuge für elektronisches Publizieren (Autorentools), Transfer, Speicherung, Konvertierung und Indexierung.
- Formale Beschreibung, Identifikation und Retrieval, Metadaten, Vernetzung.

¹⁴<http://www.global-info.org/wasistgi/index.html>, Zugriffsdatum: 20.10.2000

- Nutzung von Inhalten: Alerting, Awareness, Informationsverbünde, Informationsvermittlung, Evaluierung von Ergebnissen, Oberflächen, intelligente Agenten, Passwortproblematik.
- Wirtschaftlichkeitsmodelle, Billing (micro-billing) und Abrechnung, Statistik.

Diesen Schwerpunktbereichen stehen momentan fünf konkrete Projekte gegenüber, die im Rahmen von Global Info gefördert werden:

CARMEN Im Teilprojekt *CARMEN (Content Analysis, Retrieval and Metadata: Effective Networking)*¹⁵ werden Entwicklungen in den Bereichen Metadaten, Heterogenitätsbehandlung und Retrieval für strukturierte Dokumente mit Metadaten und heterogenen Datentypen vorgenommen. Für den Bereich Metadaten werden Werkzeuge für die Erzeugung und Verwendung von Metadaten entwickelt. Darüber hinaus sollen konzeptuelle Entwicklungen für die Domänen Terms & Conditions und Archivierung geleistet werden.

Der Bereich Heterogenitätsbehandlung soll anbieterübergreifende Suchprozesse zwischen verschiedenen Datentypen (z.B. Literaturdatenbanken und Internet) ermöglichen.

Im dritten Bereich soll eine fortgeschrittenen Retrievalkomponente auf Basis der Harvest-Architektur (siehe 4.2.5.1) entwickelt werden.

Dateninteraktives Publizieren Ziel des Projektes¹⁶ ist die Erhaltung und die Zurverfügungstellung der Primärdaten von Forschungsergebnissen in den Naturwissenschaften (insbesondere Chemie). Zur Förderung der einfachen Archivierung und Weiterverwendung wird dabei zur Zeit *CML* entwickelt, eine auf der *eXtensible Markup Language (XML)* basierende Markup-Language für Anwendungen in der Chemie.

Infrastrukturen für digitale Bibliotheken Dieses Projekt¹⁷ besteht aus den Komponenten Benachrichtigungsdienst, Brokerage Facility for Remote Access in Distributed Global Info Services und Förderierungsdienst.

Ein Benachrichtigungsdienst¹⁸ soll anbieterunabhängig einen Benutzer über neue Materialien informieren, die seinem Interessenprofil entsprechen.

¹⁵<http://www.mathematik.uni-osnabrueck.de/projects/carmen/carmen.html>, Zugriffsdatum: 20.10.2000

¹⁶<http://www.data-interaction.de/>; Zugriffsdatum: 20.10.2000

¹⁷<http://www.global-info.org/sfm/infra-kurz.htm>, Zugriffsdatum: 20.10.2000

¹⁸<http://www.inf.fu-berlin.de/~ag-db/projects/alerting/>, Zugriffsdatum: 20.10.2000

Die Komponente „Brokerage Facility“¹⁹ dient zur Entwicklung von Unterstützungsdiensten zur Auffindung, Koordination und Integration von Diensteanbietern (z.B. Verlagen, Bibliotheken, Abrechnungsdiensten, Datenbanken etc.) in digitalen, verteilten Bibliothekssystemen. Damit soll erreicht werden, dass zum einen den Diensteanbietern ein größerer Nutzerkreis geöffnet wird, und zum anderen, dass die Nutzer dieser Dienste mittels effizienter Werkzeuge den optimalen Anbieter herausfinden können.

Der Förderierungsdienst²⁰ soll Konzepte und Werkzeuge zur Literatur- und Informationsrecherche in heterogenen Informationssystemen entwickeln.

E-Verlage „Erprobung elektronischer Angebotsformen, Abrechnungsmodelle und Zahlungsverfahren auf einer Testplattform für Global Info“²¹: Ziel ist die Entwicklung eines Systems, mit dem die Tragfähigkeit einer digitalen Bibliothek evaluiert werden kann. Ausgangspunkt sind dabei die Ergebnisse und Erfahrungen der Projekte MeDOC und Interdoc. Die dabei entwickelte Testplattform soll den anderen Global Info Teilprojekten zur Erprobung der entwickelten Komponenten offenstehen.

WEP Im Rahmen dieses Teilprojekts²² werden Werkzeuge und Richtlinien für die Publikationskette, i.e. die Erstellung von Dokumenten, entwickelt. Schwerpunkte liegen auf der Entwicklung benutzerfreundlicher Autorensysteme für multimediale Publikationen, der Optimierung des Workflows in der Publikationskette und Richtlinien zur Gestaltung von Publikationen.

3.1.3 Kosten für elektronische Publikationen

Nach Berechnungen von King und Tenopir (1998) belaufen sich die minimalen Kosten für eine gedruckte wissenschaftliche Fachzeitschrift mit 500 Abonnenten auf 847 US\$ pro Subskription und Jahr. Da die Fixkosten für die Produktion den Hauptanteil der Kosten ausmachen, bedeuten mehr Abonnenten eine Kostendeckung schon bei niedrigeren Preisen. Der Wegfall der Kosten für die Reproduktion und Distribution bei einer reinen elektronischen Zeitschrift führt (bei dem Beispiel mit den 500 Abonnenten) zu einer Verringerung der Abonnementkosten um 66 US\$, wobei die zusätzlichen Kosten

¹⁹<http://vsys-www.informatik.uni-hamburg.de/projects/GlobalInfo/>,
Zugriffsdatum: 20.10.2000

²⁰http://www.witi.cs.uni-magdeburg.de/iti_db/forschung/globalinfo/,
Zugriffsdatum: 20.10.2000

²¹<http://www.e-verlage.de>, Zugriffsdatum: 20.10.2000

²²<http://www.global-info.org/sfm/wep.htm>, Zugriffsdatum: 20.10.2000

des elektronischen Mediums wie Speicherplatz, Software, Netzwerkinfrastruktur etc. noch nicht berücksichtigt sind.

Harnad (vgl. Okerson und O'Donnel, 1995) schätzt die Einsparungen bei einem Übergang zu rein elektronischen Publikationen allerdings auf 75% der Kosten für eine gedruckte Fachzeitschrift. Dies setzt allerdings ein Reengineering des Produktionsprozesses für eine Fachzeitschrift voraus (Varian, 1997).

3.2 Vorteile der elektronischen Verfügbarkeit von Dokumenten

3.2.1 Verfügbarkeit

Der offensichtlichste Vorteil von elektronisch publizierten Dokumenten ist die uneingeschränkte Verfügbarkeit, unabhängig von zeitlichen oder räumlichen Restriktionen. Ein Wissenschaftler ist weder an die Öffnungszeiten einer Bibliothek gebunden, noch muss sie/er den Weg dorthin auf sich nehmen; der Artikel in elektronischer Form kann über das Internet auf der Workstation im Büro oder zuhause gelesen werden.

3.2.2 Schnellere Publikation

Abgesehen vom Zeitgewinn bei der Beschaffung ist auch das Publizieren elektronischer Aufsätze schneller, da die Zeiten für Druck und physische Auslieferung zum Kunden wegfallen. Wiley Interscience, das elektronische Angebot von Wiley & Sons, bietet für das *Journal of the American Society for Information Science* den Online-Zugriff auf Artikel (und ganze Ausgaben der Zeitschrift), die gedruckt erst Monate später erscheinen²³.

3.2.3 Platzbedarf

Eine weitere Komponente ist der Platzbedarf in Bibliotheken und die damit verbundenen Kosten für die Aufbewahrung von Printpublikationen: Varian (1997) schätzt hier die Kosten pro Jahrgang einer Fachzeitschrift auf 35 US\$. Eine große Menge von elektronischen Publikationen kann hingegen auf billigen Massenspeichermedien vorgehalten werden. In Ginsparg (1994) gibt es hierzu ein Rechenexempel: Ein durchschnittlicher Artikel (komprimiert, mit Abbildungen) benötigt ca. 40KByte, eine (1994) handelsübliche 1 GByte Festplatte kann somit ungefähr 25.000 Artikel speichern. Bei einem Preis von 600US\$ für eine Festplatte ergibt sich somit ein Preis von

²³Mitte August 2000 war bereits der Volltext der November-Ausgabe im Netz verfügbar.

3 Cent pro Artikel²⁴.

3.2.4 Recherchemöglichkeiten

In der Welt der papierbasierten Fachzeitschriften war eine Recherche nach bestimmten Aufsätzen nur über teure Fachinformationsdienste möglich. Die elektronische Verfügbarkeit von Fachzeitschriften und Aufsätzen kann eine Formalerschließung obsolet machen, sofern für die auf Dokumentenservern abgelegten Aufsätze Metadaten verfügbar sind, die von Suchmaschinen erfasst werden können. Diese Metadatenbestände können dann sehr einfach mithilfe von Web-Suchinterfaces durchsucht werden²⁵. Zusätzlich kann durch die Auszeichnung mit Metadaten auch ein Zusammenführen inhaltsverwandter Aufsätze durchgeführt werden.

Weitere Möglichkeiten bietet die Verlinkung von Aufsätzen: Die Literaturangaben am Ende eines Aufsatzes können dabei zum Zugriff auf die Volltexte der zitierten Artikel benutzt werden (*Reference Linking*). Umgekehrt ist es auch möglich, die Aufsätze zu finden, die den gerade vorliegenden zitieren (*Citation Linking*) oder man lässt sich alle die Aufsätze anzeigen, die eine oder mehrere Quellen gemein haben (siehe 4.4).

Neben der Verlinkung von Aufsätzen ist auch die zusätzliche Verlinkung mit verwandtem Material möglich, z.B. mit Datenbanken, die DNS-Sequenzen, Proteinstrukturen oder Fotos von Galaxien enthalten. Darüber hinaus könnten Lesern die Originaldaten des Autors zur Verfügung gestellt werden, um diese zu verifizieren (Taubes, 1996).

3.2.5 Aktualisierung von elektronischen Publikationen

Ein weiterer Effekt beim Einsatz elektronischer Publikationen ist die einfache Aktualisierung. Gerade Referenzwerke oder Regelwerke in elektronischer Form können sehr einfach auf dem neuesten Stand gehalten werden. Gedruckte Werke müssen hingegen bei Änderungen neu angeschafft, oder es müssen im Falle von Loseblattsammlungen, Seiten ausgetauscht werden (Arms, 2000, S. 5).

3.2.6 Förderung der informellen Kommunikation

Zusätzlich können für elektronische Fachzeitschriften Diskussionsforen eingesetzt werden, die den unmittelbaren Autor-Leser-Kontakt ermöglichen und den Gedanken-

²⁴Durch den rapiden Preisverfall bei Massenspeichern würde Ginspargs Rechnung heute sogar noch günstiger ausfallen. Festplatten mit 20 GByte Speicherkapazität sind heute (August 2000) für unter 300 DM zu bekommen.

²⁵Springer bietet innerhalb von LINK eine Suche nach Feldern wie Titel, Autor, Erscheinungsjahretc. an.

austausch fördern. Ein Beispiel für ein solches Forum ist der Einsatz von HyperNews im *D-Lib Magazine*²⁶ (Daly, 1998).

Publikationsmodelle, die zur Förderung der informellen Kommunikation beitragen sollen, werden in Kapitel 5 kurz vorgestellt.

3.2.7 Multimedia und personalisierte Dienste

Ein Beispiel für die Ausnutzung der multimedialen Möglichkeiten elektronischer Journals ist das *Journal of Image Guided Surgery* von Jon Wiley & Sons, das benutzer-manipulierbare 3D-Modelle zur Visualisierung einsetzt (Taubes, 1996). Ein weiteres Beispiel sind die elektronischen Artikel der *Royal Society of Chemistry*²⁷, die dreidimensionale, rotierbare Modelle von Molekülen zur Verfügung stellen, die mit Hilfe eines speziellen PlugIns im Browser angezeigt werden können (Daly, 1998).

Eine weitere Zusatzfunktion ist der Einsatz von personalisierten Diensten für ein Web-Angebot. Personalisierte Dienste bieten den ihnen bekannten Benutzern ein speziell auf deren Informationsbedarf zugeschnittenes Angebot (Daly, 1998). Das Web-Portal Yahoo! bietet seinen Nutzern neben einem personalisierten Nachrichtenangebot Zusatzfunktionen wie einen kostenfreien Email-Account, einen elektronischen Kalender und ein Adressbuch²⁸. Für die wissenschaftliche Kommunikation existieren personalisierte Benachrichtigungsdienste, die den Nutzer eines Dienstes anhand seines Profils über neue Artikel informieren (Taubes, 1996).

Daly weist allerdings daraufhin, dass die Implementierung dieser zusätzlichen Funktionen in hohem Maße vom verfügbaren Budget und Personal abhängig ist.

3.3 Problemfelder der elektronischen wissenschaftlichen Kommunikation

Der Schwerpunkt der hier beschriebenen Problemfelder liegt auf den technischen Fragestellungen des elektronischen Publizierens. Rechtliche, ökonomische und organisatorische Fragestellungen werden nur miteinbezogen, wenn sie Auswirkungen auf die Infrastruktur haben.

²⁶<http://www.dlib.org>, Zugriffsdatum: 20.10.2000

²⁷<http://www.rsc.org>, Zugriffsdatum: 06.11.2000

²⁸<http://www.yahoo.com>, Zugriffsdatum: 23.10.2000

3.3.1 Information Retrieval im Internet

Die Informationsorganisation im Internet wird heute fast ausschließlich von Computerprogrammen bewerkstelligt. Die für Menschen nicht zu bewältigenden Dokumentmengen werden von speziellen Computerprogrammen, sogenannten *Crawlers* oder *Robots*, indiziert und klassifiziert. Bei diesem Prozess werden alle Wörter, die auf einer Web-Seite vorkommen, zusammen mit der URL der Seite, in der Datenbank der Suchmaschine abgelegt. Teilweise benutzen die einzelnen Suchmaschinen auch höherentwickelte Algorithmen zur Identifikation von Schlüsselwörtern. Dennoch ist die Qualität der maschinellen Indexierung nicht vergleichbar mit einer intellektuellen Indexierung. Bei einer Suche über die automatisch erstellte Datenbank mithilfe einer Suchmaschine führt dies häufig zu einer großen Anzahl irrelevanter Treffer. Da Dokumente im HTML-Format eine nur ungenügende logische Strukturierung aufweisen, sind die automatischen Indexierer nicht in der Lage, Informationen wie den Namen des Autors oder das Publikationsdatum zuverlässig aus einem Dokument zu extrahieren. Weitere Probleme entstehen durch dynamische Dokumente, i.e. Web-Seiten, die erst bei einem Aufruf durch einen *User Agent* (z.B. Web-Browser) erzeugt werden (vgl. Lynch, 1997). Bisher noch ungelöst ist das Problem der zuverlässigen Erkennung des Inhaltes von Bildern und Graphiken. Lynch (1997) weist daraufhin, dass „*Searches capable of distinguishing clearly among nudes, marmalades and national flags are still an unrealized dream.*“. Neben der Weiterentwicklung der Algorithmen zur automatischen Indexierung bieten die in 4.2 beschriebenen deskriptiven Metadaten kurz- und mittelfristig Lösungen für diese Probleme.

Ebenso schwerwiegend wie die Suchmaschinenproblematik ist das „Verschwinden“ von Dokumenten von einem bestimmten Speicherort. Dokumente im WWW werden über ihren Speicherort - den *Uniform Resource Locator (URL)* - identifiziert. Wird ein Dokument von dem angegebenen Speicherort gelöscht, z.B. durch eine Reorganisation der Serverstruktur, ist es unter der bekannten URL nicht mehr erreichbar. Notwendig sind hier Mechanismen, die den dauerhaften Zugriff auf Dokumente im WWW unabhängig von ihrem Speicherort sicherstellen. In 4.3 werden Verfahren zur dauerhaften Identifizierung und Lokalisierung elektronischer Dokumente vorgestellt.

3.3.2 Versionskontrolle bei elektronischen Dokumenten

Ein gedruckter Aufsatz spiegelt eine bestimmte Version einer Arbeit wider, ebenso wie alle Kopien dieses Aufsatzes. Die Publikation einer neuen Version (z.B. mit Korrekturen von Schreibfehlern) geht immer einher mit einem erneuten Druck. Bei Publikatio-

nen im Internet ist die Situation eine andere: Jeder, der die entsprechenden Zugriffsrechte auf eine Datei besitzt, kann diese ändern. Daher ist es sehr leicht, überarbeitete Versionen zu veröffentlichen, was natürlich auch dazu führt, dass unterschiedliche Versionen eines Aufsatzes existieren, die nicht explizit gekennzeichnet sein müssen.

Neben inhaltlichen Unterschieden sind auch die verschiedenen Dateiformate und ihr möglicherweise unterschiedliches Erscheinungsbild zu betrachten. Da die einzelnen Formate einen unterschiedlichen Leistungsumfang besitzen, stellt sich die Frage, ob ein identischer Inhalt in unterschiedlichen Formaten das gleiche Dokument darstellt. In 4.3.1 wird das Referenzmodell der IFLA vorgestellt, das eine Definition der verschiedenen Entitäten zur Verfügung stellt. Eine Vereinfachung dieses Referenzmodells, die Grundlage für die Reichweite des Digital Object Identifier (siehe 4.3.5), wird in 4.3.5.2 beschrieben (vgl. Lynch (1996) für eine detaillierte Auseinandersetzung mit den Problemen der Versionskontrolle).

3.3.3 Zitation elektronischer Dokumente

Für gedruckte Publikationen existiert ein etabliertes System für die Zitierung, z.B. nach Name des Autors, Titel, Ausgabe einer Zeitschrift und Seitenzahlen. Auf diese Weise lässt sich eindeutig auf einen bestimmten Inhalt verweisen (Lynch, 1996).

Bei elektronischen Publikationen hingegen ist die Situation bisher noch unbefriedigend, da man ohne geeignete Identifiers nur auf den Speicherort eines Dokumentes verweisen kann und auf das Zugriffsdatum²⁹.

3.3.4 Zugang zu elektronischen Aufsätzen

Wenn ein Verlag mit seinen (elektronischen) Produkten Gewinn erzielen möchte, muss er diese verkaufen. Dies bedeutet, dass der Zugang zu den Produkten beschränkt werden muss. Neben wirtschaftlichen Gründen den Zugang zu elektronischen Dokumenten zu beschränken, existieren allerdings auch noch andere. Bestimmte Ressourcen können vertraulich sein und daher darf nur eine bestimmte Personengruppe Zugang dazu haben (z.B. Geschäftsgeheimnisse). Der Zugang zu anderen Ressourcen kann an bestimmte Bedingungen geknüpft sein, beispielsweise ist der Zugang ist nur zu Lebzeiten eines bestimmten Individuums erlaubt (vgl. Arms, 2000, S. 123).

²⁹Unter <http://is.uni-sb.de/topics/zitieren.html>, Zugriffsdatum: 02.11.2000, findet man eine Aufstellung von Quellen, die das Zitieren von Online-Ressourcen zum Inhalt haben. Im englischen Sprachraum ist Li und Crane (1996) das Standardwerk für das Zitieren elektronischer Quellen.

In 4.5 werden verschiedene technische Möglichkeiten zur Regelung des Zugangs zu elektronischen Kollektionen vorgestellt.

3.3.5 Archivierung und langfristige Speicherung elektronischer Publikationen

Sommerlad (1998) unterteilt die technischen Probleme der langfristigen Archivierung in *Media Problems*, *Hardware Obsolescence* und *Software Dependence*:

- Die physischen Speichermedien wie Disketten, Festplatten oder CD-ROMs³⁰ besitzen bei weitem nicht die Langlebigkeit von Papier oder Mikrofilmen.
- Die Entwicklung neuer Speichermedien und Lesegeräte führt dazu, dass alte Speichermedien überflüssig werden und in neuen Lesegeräten nicht mehr benutzt werden können. Ein Beispiel hierfür sind die 8"-Disketten, für die keine Laufwerke mehr existieren, die an heutigen PCs angeschlossen werden können.
- Die technische Weiterentwicklung von Formaten für die Darstellung von digitalen Dokumenten ist ein großer Widersacher der langfristigen Aufbewahrung. Der Einsatz von spezialisierten Formaten erfordert auch den Einsatz spezialisierter Software für die Anzeige dieser Formate, die - wenn sie nicht weiterentwickelt oder portiert wird - auf zukünftigen Rechnergenerationen nicht mehr lauffähig ist.

Als Lösung für dieses Problem wird bei der langfristigen Aufbewahrung elektronischer Dokumente das sogenannte *digital refreshing* angewandt, i.e. das Umkopieren von alten Datenträgern auf neue. Das Problem der Formate wird in der Norwegischen Nationalbibliothek mit normalem ASCII-Text für unformatierte Dokumente und mit dem Adobe Acrobat-Format³¹ für formatierte Texte gelöst (Solbakk).

Neben den technischen Fragestellungen ergeben sich bei der langfristigen Archivierung elektronischer Publikationen auch Probleme rechtlicher und organisatorischer Natur. So ist die Selektion des zu archivierenden Materials eine entscheidende Frage. Wann ist ein elektronisches Dokument würdig, für einen längeren Zeitraum aufbewahrt zu werden, und wer sammelt die elektronischen Publikationen? Die Deutsche Bibliothek in Frankfurt hat den gesetzlichen Auftrag, alle in Deutschland erscheinenden Publikationen zu sammeln und zu archivieren. Bisher umfasst die gesetzliche Regelung

³⁰Die Aufbewahrungsdauer von CD-ROMs wird auf ca. 100 Jahre geschätzt (Arps 1993, zitiert nach Solbakk).

³¹Viewer für das Adobe Acrobat-Format sind für fast alle Rechnerplattformen verfügbar, dennoch stellt sich die Frage, ob das auch in zwanzig Jahren noch so sein wird.

allerdings nur elektronische Publikationen auf physischen Datenträgern, an Netzpublikationen werden bisher nur elektronische Dissertationen gesammelt³².

Bei gedruckten Fachzeitschriften haben Bibliotheken die vollständige Kontrolle über die langfristige Archivierung. Bei Abonnements zu elektronischen Journals sind Universitätsbibliotheken allerdings häufig nur Zugangsanbieter und besitzen die Daten in physischer Form gar nicht. Die Rechtslage ist hier nicht eindeutig, und die Handhabung dieser Fragestellung ist abhängig von den Vereinbarungen der Bibliotheken oder der Zugangsanbieter mit den entsprechenden Verlagen (Sommerlad, 1998).

Arms (1999b) weist in einer Studie über die Archivierungspolitik von drei Verlagen auf den Einfluss der Organisationsstabilität auf die Langzeitarchivierung hin. Was passiert, wenn ein Verlag sich auflöst? Elsevier Science gibt hierzu ein klares Statement ab³³:

Elsevier Science understands the importance to its library customers of providing a secure electronic archive. Therefore, the company also makes the commitment that, in the unlikely future event that Elsevier Science (or ScienceDirect itself) should be unable to meet this responsibility, the archive will be turned over to one or more depositories chosen jointly by the publisher and an independent board of library advisors.

3.3.6 Interoperabilität

Eines der Hauptprobleme im Bereich des elektronischen Publizierens lässt sich unter dem Schlagwort *Interoperabilität* fassen. Miller (2000b) definiert die Eigenschaft *interoperable* als:

one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organisation are managed in such a way as to maximise opportunities for exchange and re-use of information, whether internally or externally

Für den Bereich der elektronischen Publikationen bedeutet dies, dass die Systeme und Organisationskulturen der Informationsanbieter eine Austauschbarkeit von Informationsgütern untereinander ermöglichen und die Fähigkeit zur Zusammenarbeit besitzen. Ein konkretes Beispiel für Interoperabilität (auf einer einfachen Ebene) stellen

³²<http://deposit.ddb.de/index.htm>, Zugriffsdatum: 16.08.2000

³³http://www.sciencedirect.com/science/page/static/scidir/static_scidir_splash_about3.html, Zugriffsdatum: 05.09.2000

die sogenannten Verbundkataloge dar, die eine Recherche über mehrere Datenbanken hinweg ermöglichen.

Interoperabilität wird von Lynch und Garcia-Molina (1995) als ein Spektrum beschrieben, das von der *superficial uniformity of navigation and access* über *syntactic interoperability* bis hin zur *deep semantic interoperability* reicht. Die einfachste Form ist auf der Nutzerebene angesiedelt und stellt systemübergreifend einheitliche Benutzerinterfaces zur Verfügung. Die syntaktische Interoperabilität ermöglicht den Austausch von Metadaten und den Einsatz durchgängiger Übertragungsprotokolle für digitale Objekte. Noch Forschungsgegenstand sind Systeme, die eine tiefe semantische Interoperabilität ermöglichen, da hier die Systeme die bereitgestellten und ausgetauschten Informationen *verstehen* müssen.

Arms (2000, S. 69-72) nennt einige Bereiche, in denen die Interoperabilität von Bedeutung ist:

User Interfaces Im Bereich der Benutzerschnittstellen bedeutet Interoperabilität, dass ein Nutzer verschiedene Kollektionen von Aufsätzen (oder auch anderen digitalen Objekten) mit einer einheitlichen Benutzer-Schnittstelle durchsuchen kann.

Naming and Identification Die Namensgebung und die Identifizierungsschemata müssen eine globale Reichweite haben.

Formats Unterschiedliche Dateiformate können Benutzern einiges an Kopfzerbrechen bereiten, da man für jedes Format ein entsprechendes Viewer-Programm benötigt oder Konvertierungsprogramme, die eine Datei von einem Format A in ein Format B umsetzen können. Die Trennung von Inhalt und Layout, wie es mit der *eXensible Markup Language (XML)* und Stylesheet-Sprachen wie der *eXtensible Style Language (XSL)* propagiert wird, könnte hier Abhilfe schaffen.

Metadata Für die Suche in verteilten Beständen oder über Kollektionen unterschiedlicher Objekte muss eine Austauschbarkeit der Metadaten zwischen den Systemen möglich sein. Die Austauschbarkeit muss sowohl hinsichtlich der Syntax der Metadaten als auch der Semantik der Metadaten-Attribute hergestellt sein. In 4.2 wird mit dem *Resource Description Framework* eine Syntax für Metadaten vorgestellt, semantische Interoperabilität kann durch *Metadata Registries* und *Namespaces* erreicht werden. Möglichkeiten zur Herstellung von Interoperabilität auf Metadatenebene werden in 4.2.3 vorgestellt.

Distributed Searching Die Suche in verteilten Datenbeständen ist ein komplexes Problem, das unter anderem die Bereiche Metadaten und Suchprotokolle beinhaltet. Im einfachsten Fall kann die Interoperabilität zwischen verschiedenen Kollektionen durch den Einsatz von standardisierten Metadatensets und die Verwendung gleicher Suchprotokolle hergestellt werden. Da dies aber unrealistisch ist, müssen Mechanismen gefunden werden, die eine Suche auch über unterschiedlich organisierte Datenbestände erlauben.

Network Protocols Die weitgehende Akzeptanz des IP-Protokolls ermöglicht schon heute die Kommunikation zwischen unterschiedlichsten Systemen.

Retrieval Protocols Ein ideales Retrievalprotokoll würde die sichere Authentifizierung zweier Computer, Abfragen, welche Datenbestände jeder Rechner bietet, eine Vielzahl an Suchmöglichkeiten, Methoden zur Speicherung und Modifizierung von Zwischenresultaten und Schnittstellen zu vielen Formaten und Protokollen bereitstellen.

Ein Versuch, Interoperabilität zwischen verschiedenen Bibliothekskatalogen und damit auch die simultane Suche über mehrere Kataloge hinweg zu erreichen, ist das Z39.50 Protokoll. Miller (1999) beschreibt einige beispielhafte Anwendungen für Z39.50.

Authentication and security Eine Authentifizierung muss auf drei Ebenen stattfinden: Nutzer müssen authentifiziert werden, Rechnersysteme müssen authentifiziert werden, damit bei sicherheitskritischen Anfragen festgestellt werden kann, welcher Rechner anfragt, und zuletzt müssen auch die Objekte einer digitalen Sammlung authentifiziert werden, damit ein Nutzer sicher sein kann, dass das angeforderte Objekt nicht modifiziert wurde.

In Miller (2000b) werden neben den technischen Aspekten weitere Bereiche beschrieben, in denen Interoperabilität notwendig ist. Die Entscheidung für eine Interoperabilität auf der technischen Seite mit einer Öffnung der eigenen Datenbestände nach außen, hat Implikationen für die Struktur einer Organisation, da zum einen dem Personal neue technische Fertigkeiten abverlangt werden (bzw. das Lernen neuer Fertigkeiten), und zum anderen sieht sich das Personal plötzlich einem größeren und womöglich anders gearteten Nutzerkreis gegenüber. Neben diesen organisatorischen Erfordernissen ist eine Interoperabilität auch hinsichtlich verschiedener Disziplinen nötig. Bibliotheken, Verwaltungen und Museen sehen sich teilweise ähnlichen Fragestellungen gegenüber, denen mit generalisierten Lösungen begegnet werden kann.

Der Datenaustausch zwischen Organisationen ist häufig auch beschränkt durch rechtliche Restriktionen, die z.B. aufgrund der Datenschutzbestimmungen eine Weitergabe von Daten an andere Organisationen verhindern. In Deutschland sind die Krebsregister der Länder mit ihren Beschränkungen hinsichtlich der Datensammlung und Datenweitergabe zur Forschung ein gutes Beispiel (vgl. Pommerening, 1993).

Die Schwierigkeit der oben beschriebenen Fragestellungen erhält noch eine zusätzliche Komplexität, wenn nationale Grenzen und vor allem auch Sprachgrenzen überschritten werden (vgl. Baker, 1998).

3.4 Ein Szenario für die formelle wissenschaftliche Kommunikation über das Internet

Das folgende Szenario geht von einem akademischen Nutzer an einer Universität oder einem Forschungsinstitut aus. Der Publikationsprozess stellt eine Mischform zwischen dem *Modernized Garvey/Griffith Model* der wissenschaftlichen Kommunikation von Hurd (1996, S. 22) und dem *No-Journal Model* von Hurd (1996, S. 25) dar. Abbildung 3.1 stellt den Publikationsprozess schematisch dar. Das Modell favorisiert kein bestimmtes Geschäftsmodell, d.h. kostenfreie Dokumentenserver und e-Journals können mit kostenpflichtigen Angeboten von Verlagen koexistieren.

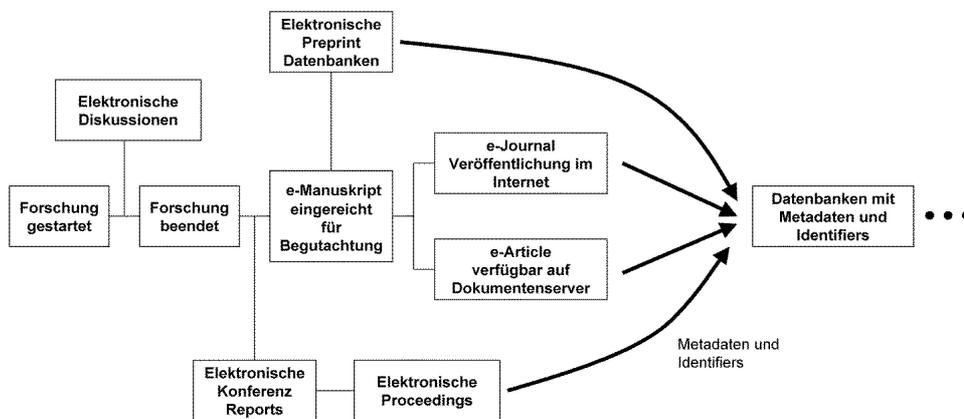


Abbildung 3.1: Publikationsprozess mit einem Nebeneinander von elektronischen Zeitschriften und Dokumentenservern

Der Forschungsprozess wird begleitet von informellen elektronischen Diskussionen; die aktuellen Forschungsarbeiten werden auf Tagungen vorgestellt und in elektronischen Tagungsbänden publiziert. Wissenschaftliche Ergebnisse werden in Form von Artikeln auf Preprint-Servern veröffentlicht. Parallel dazu werden sie einem Peer Review unterzogen und entweder in einer elektronischen Zeitschrift oder einem durch-

suchbaren Dokumentenserver abgelegt. Jeder Autor stellt die Metadaten für seinen Aufsatz in einem geeigneten Format zur Verfügung. Zusätzlich enthält jeder veröffentlichte Aufsatz einen eindeutigen Identifier mit dem Aufsätze unabhängig von ihrem Speicherort gefunden werden können. Der Volltext der Artikel und die Metadaten inklusive Identifiers werden von den Betreibern der Dokumentenservern an Suchmaschinenbetreiber gesendet, die die Daten aufbereiten und einem großen Nutzerkreis zur Verfügung stellen. Zusätzliche Dienste der Suchmaschinenbetreiber sind die Verlinkung von Artikeln untereinander, die Verlinkung von Artikeln mit Datenbanken, die zusätzliche Materialien zur Verfügung stellen oder auch ein Benachrichtigungsdienst für neue Artikel.

Aus der Sicht des Nutzers stellt sich diese Infrastruktur folgendermaßen dar: Mit dem Suchformular auf der Website eines der Suchmaschinenbetreiber kann man gezielt die Metadaten der erfassten Dokumente durchsuchen (z.B. Suche nach Autor „Schmitt“ und Jahr „2000“). Als Suchergebnis erhält man eine Liste mit Identifiers der gefundenen Aufsätze. Ein Klick auf einen der Treffer stößt einen Prozess an, der den Identifier zu einer oder mehreren URLs auflöst. Falls mehrere URLs angeboten werden, kann man auswählen, ob man beispielsweise das Inhaltsverzeichnis, ein Abstract, das Literaturverzeichnis oder den Volltext eines Aufsatzes möchte. Abhängig von den Lizenzvereinbarungen der Institution ist der Zugriff auf den Volltext eines Artikels möglich oder auch nicht möglich; im zweiten Fall wird stattdessen nur das Abstract oder Inhaltsverzeichnis des gewünschten Artikels angezeigt, oder aber man kann einzeln für diesen Artikel über geeignete Micropayment-Verfahren bezahlen (pay-per-view). Alle in einem Aufsatz zitierten Quellen, die in elektronischer Form verfügbar sind und von dem Suchmaschinenbetreiber in seiner Datenbank abgelegt sind, können über in die Literaturangaben eingebettete Links erreicht werden. Zusätzlich können für den gerade vorliegenden Aufsatz alle Zitationen angezeigt werden. Eine weitere Funktion des Suchmaschinenbetreibers für den Nutzer ist die Möglichkeit der Abspeicherung eines Interessenprofils für einen Benachrichtigungsdienst. So kann man per Email über neue Artikel in der Datenbank des Suchmaschinenbetreibers informiert werden, deren Metadaten Übereinstimmungen mit dem abgespeicherten Interessenprofil aufweisen.

Kapitel 4

Infrastruktur für die wissenschaftliche Kommunikation über das Internet

In diesem Kapitel werden existierende Technologien für eine Infrastruktur beschrieben, mit dessen Hilfe sich das in 3.4 vorgestellte Szenario für die wissenschaftliche Kommunikation über das Internet umsetzen lässt.

4.1 Objekte und Dokumente

In der Welt der papierbasierten wissenschaftlichen Kommunikation ist die höchste Aggregationsstufe für den Austausch von Informationen ein Paket mit mehreren Dokumenten, repräsentiert z.B. durch eine Fachzeitschrift mit ihren Ausgaben und den darin enthaltenen Aufsätzen oder ein Buch mit Kapiteln von mehreren Autoren. Die nächsttiefere Stufe ist dann das *Dokument*, üblicherweise ein wissenschaftlicher Aufsatz in einer Fachzeitschrift oder ein Buchkapitel. Allerdings gibt es auch andere Manifestationen, die unter *Dokument* fallen, z.B. eine geographische Karte. In der elektronischen Kommunikation gibt es nun eine weitere Untergliederungsstufe, da einzelne Komponenten eines Dokumentes identifiziert werden müssen. In einer Hypertextumgebung wie dem World Wide Web besteht ein Aufsatz oftmals aus mehreren miteinander verlinkten Dateien. Ebenso kann eine einzelne Web-Seite aus mehreren Einzeldateien bestehen (Text, Abbildungen, Audio etc.) (vgl. Paskin, 1997).

Arms (1995) schlägt für die Informationen, die in digitalen Bibliotheken abgelegt werden, das Konzept des *digital object* vor. Ein digitales Objekt besteht aus zwei Teilen: dem tatsächlichen Inhalt, der für einen Nutzer relevant ist (z.B. ein Aufsatz, eine Illustration etc.) und Metadaten, die das Objekt hinsichtlich verschiedener Aspekte be-

schreiben. Zu den Metadaten gehören ein *Handle*, i.e. Identifier mit zugehörigem Resolutionsmechanismus (vgl. 4.3), und *Properties*, i.e. Metadaten zum Resource Discovery oder auch zur Festlegung des Zugangs zu diesem Objekt. Darüber hinaus ist für verschiedene Anwendungen der Einsatz eines *Transaction log* sinnvoll, in dem alle Transaktionen in Verbindung mit dem digitalen Objekt festgehalten werden können und eine digitale Signatur, mit der die Authentizität des Objekts überprüft werden kann.

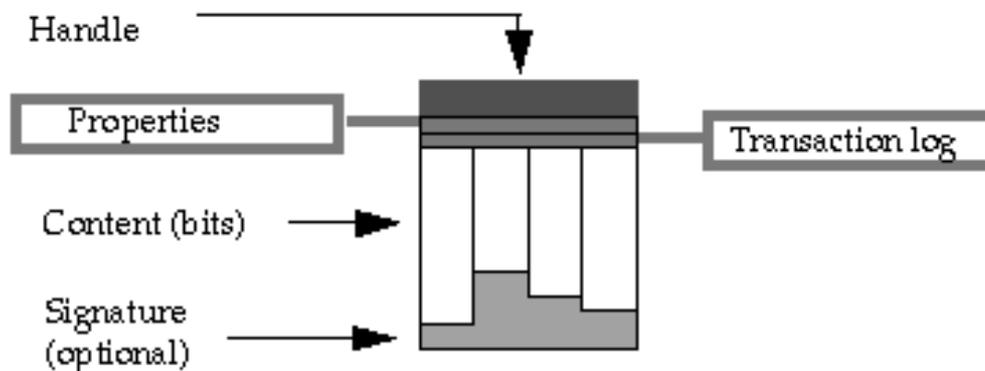


Abbildung 4.1: Teile eines digitalen Objekts (Arms, 1995)

Mithilfe sogenannter *Meta-objects* können digitale Objekte gruppiert werden. Ein Meta-object enthält die Handles der enthaltenen Objekte und Informationen über ihre Beziehungen zueinander. Ein Beispiel für ein Meta-object könnte die August-Ausgabe einer Fachzeitschrift mit ihren Artikeln sein.

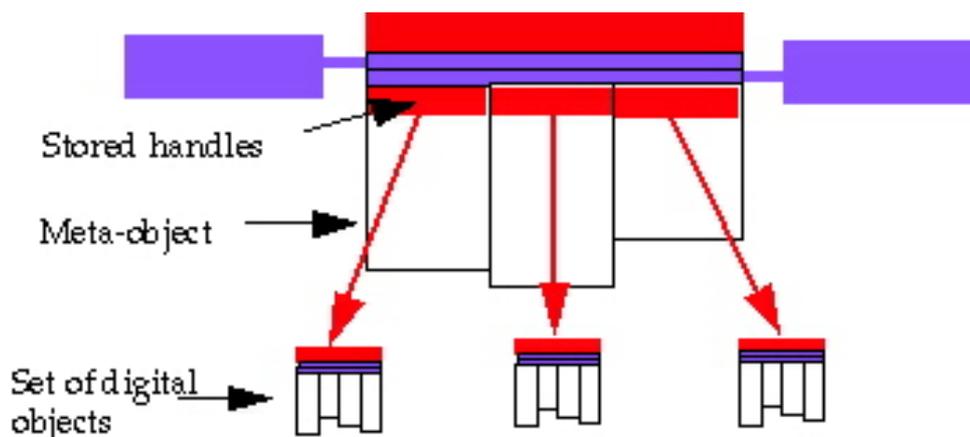


Abbildung 4.2: Meta-object (Arms, 1995)

Arms (2000, S. 230/1) weist daraufhin, dass neben statischen Objekten in einer

digitalen Bibliothek auch dynamische Objekte, wie z.B. Computerprogramme, Simulationen, Sensordaten etc. abgelegt sein können. Die Verwaltung dieser Objekte wird nachfolgend nicht berücksichtigt, da hier noch keine etablierten Verfahren existieren.

4.2 Metadaten

4.2.1 Funktion von Metadaten

Die Bedeutung des Terminus *Metadaten* ist „Daten über Daten“. Metadaten sind einem Objekt zugeordnete Informationen, die dieses hinsichtlich verschiedener Aspekte beschreiben können.

Metadaten sind kein neues Konzept - seit Jahrhunderten erschliessen Bibliotheken ihre Sammlungen auf Katalogkarten (vgl. Barth, 1996). Ein neueres Beispiel sind die Datenbankeinträge für Objekte in einem Bibliotheks-OPAC. Sowohl Katalogkarten als auch Datenbankeinträge enthalten Informationen über das Objekt, wie z.B. den Titel, den Autor, das Erscheinungsjahr, die Verlagsangabe, den Standort, Schlagwörter etc..

Bearman und Sochats (1994) schlagen ein Referenzmodell für *Business Acceptable Communications* vor, in dem sechs verschiedene Schichten von Metadaten enthalten sind:

Handle Hierunter fällt das Resource Discovery und Retrieval sowie die Zuordnung von Identifiers zu einem Objekt.

Terms & Conditions In diesen Metadaten werden z.B. Nutzungsgebühren, Zahlungsbedingungen etc. festgelegt.

Structure Hier wird die logische Struktur eines Dokumentes, das aus verschiedenen Objekten (Text, Audio, Video etc.) bestehen kann, festgehalten und es wird beschrieben, wie die verschiedenen Objekte zu dem Dokument zusammengefügt werden.

Context Hier wird die Erzeugung oder der Gebrauch eines Objektes mit einem Ereignis in einen Zusammenhang gebracht.

Content Der Inhalt, der in einer Transaktion übertragen wird.

History of Use Hier wird der Verlauf der Benutzung des Objekts festgehalten: Wie das Objekt benutzt wurde, wann es benutzt wurde und von wem.

Für eine optimierte Nutzbarkeit von elektronischen Information in der wissenschaftlichen Fachkommunikation ist bei diesem Modell insbesondere die *Handle Layer* von Interesse. Im folgenden Abschnitt wird mit dem Dublin Core Metadaten-set das prominenteste Beispiel für das Resource Discovery von Objekten im Internet vorgestellt.

Die Schichten *Terms & Conditions* und *History of Use* sind insbesondere im Bereich des Zugangsmanagements (siehe 4.5) von Bedeutung, akzeptierte Standards wie Dublin Core existieren hier allerdings noch nicht.

Die Schicht *Structure* findet sich teilweise im *Resource Description Framework* wieder, das in 4.2.4 dargestellt wird.

4.2.2 Dublin Core als ein Beispiel für deskriptive Metadaten

4.2.2.1 Bisherige Entwicklung

Aus der Notwendigkeit heraus, die Informationen aus Bibliothekskatalogen auszutauschen, entstand mit *MARC (MAchine READable Catalog)* in den 60er Jahren ein heute in verschiedenen Ausprägungen international akzeptierter Standard (Dempsey und Heery, 1997). Die Erstellung und Pflege von MARC-Einträgen erfordert Training und Erfahrung, da die Anwendung der Katalogisierungsregeln komplex ist. Die Katalogisierung von Objekten mithilfe von MARC ist für das Resource Discovery von Objekten zwar sehr effektiv, für die Beschreibung von Web-Ressourcen aber zu aufwendig. Die Datenmengen des World Wide Web sind mit einem nur für Experten anzuwendenden Metadatenformat wie MARC nicht zu bewältigen. Aus diesem Missstand heraus entwickelte sich 1995 das *Dublin Core Element Set (DCES)*, das zum einen die notwendigen Elemente besitzt, um ein Resource Discovery im Internet zu ermöglichen, zum anderen aber einfach genug ist, um eine weitreichende Verbreitung zu ermöglichen.

Die Entwicklung von Dublin Core wird getragen von der *Dublin Core Metadata Initiative (DCMI)*¹, die ihr Direktorat beim Online Computer Library Center (OCLC)² hat. Seit 1995 werden Workshops veranstaltet, die dem Austausch und der Konsensbildung der beteiligten Interessengruppen dienen. Die Weiterentwicklung von Dublin Core wird durch die Working Groups vorangetrieben, die verschiedene Aspekte von Dublin Core bearbeiten. Der Ablauf der Weiterentwicklung orientiert sich weitgehend an anderen Standardisierungsbemühungen³. Mittlerweile sind formelle Standardisie-

¹<http://purl.org/dc/> Zugriffsdatum: 22.08.2000

²<http://oclc.org/>, Zugriffsdatum: 22.08.2000

³vgl. <http://purl.org/dc/documents/dcmi-structure>, Zugriffsdatum: 22.08.2000

rungsbestrebungen bei der CEN (European information industry standardization forum) und der NISO (the North American information standardization organization) im Gange.

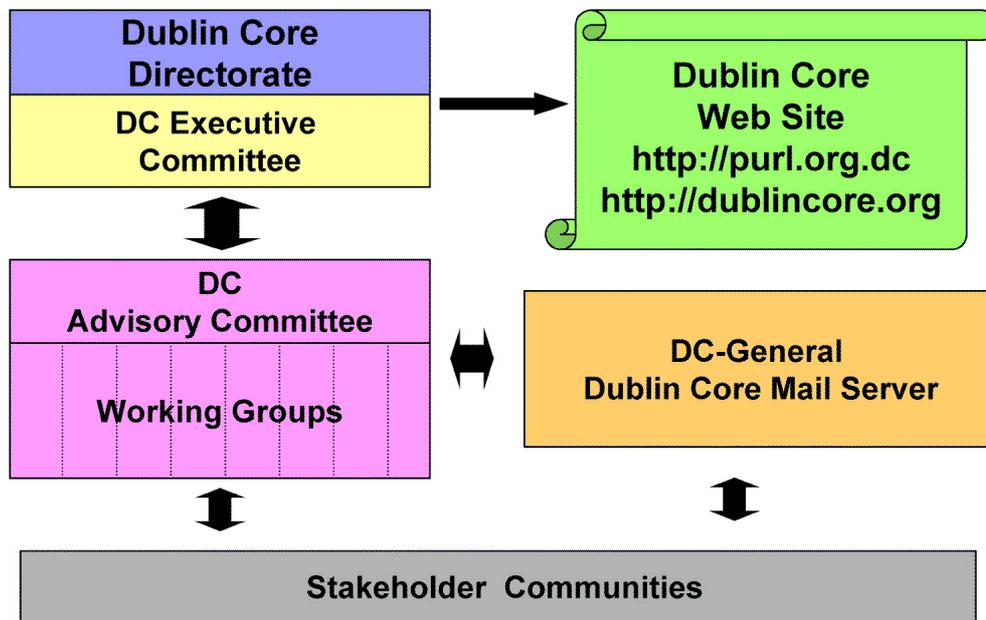


Abbildung 4.3: Organisationale Struktur der Dublin Core Metadata Initiative (Weibel, 2000b)

Dublin Core wurde ursprünglich entwickelt, um das Resource Discovery von sogenannten *document-like objects (DLOs)* zu ermöglichen. Das Konzept der DLOs wird nicht klar definiert, sondern anhand von Beispielen verdeutlicht: Die elektronische Ausgabe einer Tageszeitung ist ein DLO, eine Sammlung von unkommentierten Folien eines Vortrags hingegen nicht. DLOs sind hinsichtlich ihrer Komplexität nicht beschränkt, allerdings besteht der primäre Inhalt aus Text, und die benötigten Metadaten weisen eine starke Ähnlichkeit mit Metadaten auf, die für die Beschreibung von gedruckten Texten verwendet werden (Weibel, 1995). Seit dem dritten Dublin Core Workshop in Dublin, Ohio wird der Einsatz von DC auch auf andere Domänen, wie die Beschreibung von Bildern, propagiert. Diese Entwicklung hatte Implikationen für das Element-Set: Das Element *SUBJECT and DESCRIPTION* wurde in zwei Elemente aufgespalten (*SUBJECT* und *DESCRIPTION*); zusätzlich wurde ein Element *RIGHTS MANAGEMENT* aufgenommen (Weibel und Miller, 1997). Die Domäne Bildbeschreibung illustriert auch recht gut die Probleme des *1:1-Prinzips* für die Ressourcenbeschreibung. Dieses Prinzip sagt aus, dass jeder Ressource genau ein Metadaten-Satz zugeordnet wird. Welcher Wert wird beispielsweise dem Attribut *CREATOR* bei ei-

nem digitalisierten Foto eines Gemäldes von Matisse zugeordnet? Der Name desjenigen, der das Bild eingescannt hat oder der Name des Künstlers? Mithilfe des RELATION-Elements können solche Beziehungen zwar ausgedrückt werden, dennoch ist das 1:1-Prinzip gegenwärtig in der Diskussion (vgl. DC-FAQ, 2000).

Aus der Diskussion um die Reichweite von Dublin Core haben sich vier Hauptanforderungen herauskristallisiert, die die weitere Entwicklung von Dublin Core leiten (Weibel und Lagoze, 1997):

Simplicity of Creation and Maintenance Eine Beschreibungssprache für Objekte im World Wide Web muss einfach sein, damit sie auch ein in der Beschreibung von Objekten ungeübter Autor einsetzen kann. Die Qualität der Metadaten erreicht dann zwar nicht die Qualität von MARC-Beschreibungen, dennoch ermöglicht sie ein besseres Resource Discovery als die Volltext-Indexe der Suchmaschinen.

Commonly understood Semantics Der Einsatz einer domänen-unabhängigen Beschreibungssprache mit einer Beschränkung auf wenige, allgemein verständliche Kern-Elemente ermöglicht ein disziplinenübergreifendes Finden von Informationen.

International Scope Für eine globale Informationsinfrastruktur ist die Berücksichtigung der verschiedenen Sprachen unerlässlich. Gegenwärtig existieren mehrere Übersetzungen des Dublin Core Metadaten Sets, u.a. in Arabisch, Chinesisch, Niederländisch, Finnisch, Französisch, Deutsch⁴, Griechisch, Japanisch, Koreanisch, Norwegisch, Portugiesisch, Spanisch, Schwedisch und Thai (vgl. Baker, 1998).

Extensibility Einerseits ist für die einfache Handhabung und Akzeptanz des DCES eine Beschränkung der Anzahl der Beschreibungselemente notwendig, andererseits benötigen bestimmte Anwendungen mehr Attribute für die Beschreibung von Objekten.

4.2.2.2 Unqualifiziertes vs. qualifiziertes Dublin Core

Das Dublin Core Element Set besteht heute aus 15 Elementen (vgl. Tabelle 4.1), die einen „Kern“ für die Beschreibung von Objekten bilden, d.h. es sind die Elemente aufgeführt, die unabhängig von einer spezifischen Domäne eine sinnvolle Beschreibung

⁴Die deutsche Übersetzung von Rusch-Feja ist unter <http://rdf.dev.oclc.org/eor/2000/03/13-dces.de> verfügbar (Zugriffsdatum: 13.11.2000).

ermöglichen. Dabei sind allerdings Spezialanforderungen bestimmter Domänen noch nicht berücksichtigt. Baker (1998) zieht eine Analogie von Dublin Core zu Pidgin Sprachen. In der realen Welt entstanden Pidgin Sprachen mit kleinem Wortschatz, geringer Flexion und freier Wortfolge aus der Notwendigkeit heraus, dass Menschen mit unterschiedlicher Muttersprache miteinander arbeiten oder Handel treiben mussten. Eine ähnliche Situation hat sich mit der Verbreitung des World Wide Web für die unterschiedlichsten Fachdisziplinen ergeben: Plötzlich hatten sie über das WWW die Möglichkeit, auf einfache und schnelle Art und Weise zusammenzuarbeiten und Ergebnisse auszutauschen. Dublin Core ist die Pidgin Sprache für die Fachgemeinschaften.

<i>Name</i>	<i>Identifier</i>	<i>Definition</i>
TITLE	TITLE	A name given to the resource
CREATOR	CREATOR	An entity primarily responsible for making the content of the resource.
SUBJECT AND KEY-WORDS	SUBJECT	The topic of the content of the resource.
DESCRIPTION	DESCRIPTION	An account of the content of the resource.
PUBLISHER	PUBLISHER	An entity responsible for making the resource available.
CONTRIBUTOR	CONTRIBUTOR	An entity responsible for making contributions to the content of the resource.
DATE	DATE	A date associated with an event in the life cycle of the resource.
RESOURCE TYPE	TYPE	The nature or genre of the content of the resource.
FORMAT	FORMAT	The physical or digital manifestation of the resource.
RESOURCE IDENTIFIER	IDENTIFIER	An unambiguous reference to the resource within a given context.
SOURCE	SOURCE	A Reference to a resource from which the present resource is derived.
LANGUAGE	LANGUAGE	A language of the intellectual content of the resource.
RELATION	RELATION	A reference to a related resource.
COVERAGE	COVERAGE	The extent or scope of the content of the resource.
RIGHTS MANAGEMENT	RIGHTS	Information about rights held in and over the resource.

Tabelle 4.1: Das Dublin Core Element Set, Version 1.1 (REC-DCES, 1999)

Der Terminus „*unqualifiziertes*“ *Dublin Core* verweist dabei auf die alleinige Benutzung dieser 15 Elemente, die als Attribut-Wert-Paare dargestellt werden.

In der Anwender- bzw. Entwicklergemeinde von *Dublin Core* gibt es unterschiedliche Ansichten darüber, wie das DCES eingesetzt werden sollte. Das eine Extrem stellen die sogenannten Minimalisten dar, die die 15 Kernelemente für die Beschreibung von Ressourcen als ausreichend ansehen, das andere Extrem sind die sogenannten Strukturalisten, die zusätzlichen Kontext und die Qualifikation von Attributen als notwendig erachten (Weibel, 1997). Gegenwärtig werden zwei Klassen von Qualifiers unterschieden (DCMES-QUALIFIERS, 2000):

Element Refinement Diese Art von Qualifier spezifiziert ein Element genauer, wobei die Bedeutung des unqualifizierten Elementes erhalten bleibt.

Für das DC-Element DATE werden z.B. die Qualifiers *Created*, *Valid*, *Available*, *Issued* und *Modified* empfohlen (vgl. DCMES-QUALIFIERS, 2000). Auf eine Web-Seite bezogen führt der Einsatz der Qualifiers *Created* und *Modified* zu mehr Klarheit bezüglich der Aktualität der dargebotenen Informationen.

Encoding Scheme Hier wird ein *Scheme* spezifiziert, das die Interpretation des Element-Wertes unterstützt. Schemes können kontrollierte Vokabularien (z.B. Dewey Decimal Classification, Library of Congress Subject Headings) sein, formale Notationen (z.B. Datum nach ISO-Norm) oder Parsing-Regeln.

Der Wert eines SUBJECT-Attributes kann z.B. mit dem Kodierungsschema LCC *qualified* sein, d.h., dass der Wert des Attributes SUBJECT in der *Library of Congress Classification*⁵ enthalten ist.

Ein wesentliche Anforderung an die Qualifiers ist, dass sie die Grundbedeutung des ursprünglichen Elementes nicht verändern, d.h. das qualifizierte Element muss auch Sinn machen, ohne dass die Bedeutung des Qualifiers verstanden wird. Der Prozess der Rückwandlung zur unqualifizierten Form wird als *dumbing-down* bezeichnet. Die Qualifikation von *Dublin Core* Attributen ermöglicht zwar die präzisere Beschreibung von Ressourcen, bedeutet aber auch eine größere Komplexität, die die Interoperabilität zwischen verschiedenen Fachgemeinschaften reduzieren kann. Abbildung 4.4 zeigt, wie das *dumbing-down* bei Einsatz eines Qualifier⁶ zu ungewollten Suchergebnissen führen kann. Beim DCMI wird als Lösung für dieses Problem ein *usage board* vorge-

⁵<http://lcweb.loc.gov/catdir/cpsol/lcco/lcco.html>, Zugriffsdatum: 29.08.2000

⁶Der Qualifier *Affiliation* wurde zwar vorgeschlagen, aus den beschriebenen Gründen aber nicht zugelassen (Lagoze, 2000).

schlagen, dass die Nutzung von Qualifiers in verschiedenen Fachgemeinschaften überwacht und eine *registry* mit Qualifiers pflegt, in der die Interoperabilität festgehalten ist (Lagoze, 2000).



```
<meta name="creator" content="Alison Lurie">
<meta name="creator:affiliation" content="Cornell">
```



```
<meta name="creator" content="Gary Cornell">
```

Find me objects of which Cornell is the creator



Abbildung 4.4: Unkontrollierte Qualifikation vs. Interoperabilität (Lagoze, 2000)

4.2.2.3 Speicherung von Dublin Core Metadaten

Dublin Core Metadaten können auf verschiedene Weise abgelegt werden. Für HTML-Dokumente bietet sich die direkte Kodierung in den Quelltext an: Mithilfe des `link`-Tags an wird das einzubindende Dublin Core Schema mit einer URL angegeben (hier das Dublin Core Metadata Element Set Version 1.1); `meta`-Tags nehmen die Attribut-Wert-Paare auf. Mit dem Präfix `DC` wird ausgedrückt, dass die verwendeten Attribute aus dem Dublin Core Element Set stammen. Folgendes Beispiel zeigt die Kodierung von Dublin Core Metadaten für eine mögliche HTML-Version dieser Arbeit:

```
<link rel = "schema.DC"
href = "http://purl.org/DC/elements/1.1/">
<meta name = "DC.Title"
content = "Optimierung der informationellen
Infrastruktur für die wissenschaftliche Kommunikation">
<meta name = "DC.Creator" content="Weiland, Peter">
<meta name = "DC.Date" content = "2000">
<meta name = "DC.Format" content = "text/html">
<meta name = "DC.Language" content = "de">
```

Mittlerweile existieren schon einige Suchmaschinen, die in HTML eingebettete unqualifizierte Dublin Core Metadaten auswerten können (siehe 4.2.5.1), Werkzeuge für die Auswertung von qualifiziertem Dublin Core in HTML sind hingegen noch nicht verbreitet (Cox u. a., 2000). Ein Manko der Einbettung von Dublin Core in HTML ist die fehlende Ausdrucksmöglichkeit für die Gruppierung von Elementen und Strukturierung (vgl. DC-FAQ, 2000). Für bestimmte Anwendungen kann diese Einschränkung durch die Wiederholung von Attributen mit anderen Werten bzw. die Zuweisung von Listen von Werten zu einem Attribut umgangen werden (Cox u. a., 2000, Abschnitt 3.3).

Die beste Unterstützung für die Funktionen *Suche* und *Retrieval* bietet das Ablegen der Dublin Core Metadaten in einem Database Management System (DBMS). Da hier eine Trennung des Objekts von seinen Metadaten vorgenommen wird, verdient die Einhaltung der Konsistenz und die Synchronisation besonderes Augenmerk (DC-FAQ, 2000).

Die leistungsfähigste Art der Kodierung von Dublin Core Metadaten bietet das vom *World Wide Web Consortium (W3C)*⁷ veröffentlichte *Resource Description Framework (RDF)*. RDF wird in 4.2.4 genauer beschrieben.

4.2.2.4 Das Nordic Metadata Project als Beispiel für eine Implementierung

Das *Nordic Metadata Project* (Hakala u. a., 1998), das von Oktober 1996 bis Juni 1998 lief, war eines der ersten Projekte, das Dublin Core zur Beschreibung elektronischer Ressourcen nutzte. Ziel des Projekts war die Schaffung einer Umgebung für die Produktion, Indexierung und das Retrieval von Metadaten, wobei die entwickelten Werkzeuge tatsächlich im Produktivbetrieb eingesetzt werden sollten und nicht nur Testzwecken dienen sollten. Folgende Werkzeuge wurden im Rahmen des Projekts entwickelt:

⁷<http://www.w3.org>, Zugriffsdatum: 02.11.2000

Metadata template Das *metadata template*⁸ ist ein interaktives, webbasiertes Werkzeug zur einfachen Erstellung von Dublin Core Metadaten.

Metadata harvesting and indexing application Da im Projekt keine Ressourcen für die Neuentwicklung einer Suchmaschine vorhanden war, wurde die nicht-kommerzielle Nordic Web Index Software entsprechend angepasst⁹.

Dublin Core to MARC converter Dieser Konverter¹⁰ ermöglicht die Extraktion der Dublin Core Metadaten eines Dokumentes und wandelt diese in das MARC-Format um.

URN generator Diese Komponente¹¹ ermöglicht die Vergabe von Uniform Resource Names (URN), d.h. eindeutigen Identifiers für abgelegte Ressourcen. In 4.3.4.5 wird der URN-Generator des Nordic Metadata Project kurz vorgestellt.

Neben den entwickelten Werkzeugen ist als Erfolg des Projekts die Existenz von zwei nationalen Metadaten-Datenbanken zu verzeichnen: *Swemeta* und *Danmeta*. Beide Datenbanken enthalten ungefähr 250.000 Einträge, dies entspricht 6-7% des gesamten Webspaces in diesen Ländern¹².

Im Januar 1999 wurde das *Nordic Metadata II Project* gestartet, das neben der Fortführung der Projektinhalte des ersten Projekts, vor allem die Verbesserung der Konvertierung von Dublin Core nach MARC und die Einbeziehung neuer Syntaxstrukturen wie RDF/XML und XHTML¹³ in die vorhandenen Werkzeuge vorsieht.

4.2.3 Interoperabilität auf Metadaten-Ebene

4.2.3.1 Schemas

Neben Dublin Core als ein Metdaten-Schema für die Beschreibung von Web-Ressourcen jeglicher Art, benötigen spezifische Anwendungen Beschreibungsmöglichkeiten, die über die Möglichkeiten der 15 Dublin Core Elemente hinausgehen.

Ein Beispiel für ein Schema, das der Beschreibung spezieller Ressourcen dient, ist das *DC Education Schema* (Mason und Sutton, 2000). DC Education besteht aus den

⁸http://www.lub.lu.se/metadata/DC_creator.html, Zugriffsdatum: 05.09.2000

⁹Das Angebot des Nordic Web Index wurde mittlerweile wegen fehlender Förderung eingestellt (siehe <http://nwi.ub2.lu.se/?lang=en>, Zugriffsdatum: 03.11.2000)

¹⁰<http://www.bibsys.no/meta/d2m/>, Zugriffsdatum: 03.11.2000

¹¹<http://www.lub.lu.se/cgi-bin/nmurn.pl>, Zugriffsdatum: 16.10.2000

¹²Nordic Metadata project - Homepage (<http://renki.lib.helsinki.fi/meta/>, Zugriffsdatum: 05.09.2000)

¹³XHTML (eXtensible Hypertext Markup Language) ist eine Neuformulierung von HTML in XML.

Standard Dublin Core Elementen, den beiden zusätzlichen Elementen *Audience* und *Standards* und den Elementen *InteractivityType*, *InteractivityLevel* und *TypicalLearningTime*. Die drei letztgenannten Elemente sind dem *IEEE Learning Object Metadata (LOM)* Standard entlehnt.

Ein Schema, das speziell auf den elektronischen Handel und die Verwaltung von Urheberrechten abzielt, ist das *indecs* Metadaten-Schema, das vom *Digital Object Identifier* (siehe 4.3.5) benutzt wird.

4.2.3.2 Registries und Crosswalks

Die Interoperabilität verschiedener Schemas ist ein bedeutender Faktor für den Austausch von Metadaten über Disziplinengrenzen hinweg. Für eine Integration der Dienste verschiedener Anbieter über das Internet, d.h. den Austausch von Metadatenbeständen und die konsistente Suche in ihnen, ist die Schaffung von sogenannten *registries* eine Lösung. Heery u. a. (2000) definiert *Metadata registries* als

[...] formal systems that can disclose authoritative information about the semantics and structure of the data elements that are included within a particular metadata scheme. Registries would typically define the semantics of metadata elements, give information on any local extensions in use, and provide mappings to other metadata schemes.

Diese Registries sind Datenbanken, in denen die Elemente domänen-spezifischer Schemas mit Richtlinien zu ihrem richtigen Einsatz abgelegt sind. Idealerweise sind diese Registries so angelegt, dass sie sowohl von Menschen als auch von Softwareprogrammen benutzt werden können. Dies ermöglicht zum einen Autoren, ihre Beschreibungen von Objekten an gängigen Standards zu orientieren, bzw. wenn für eine Domäne noch kein Standard existiert, einen eigenen zu registrieren. Zum anderen können Softwareprogramme auf die Registry zugreifen. Dies können Programme zur Manipulation von Metadaten sein, die daher auf die Syntax und Semantik eines Schemas zugreifen müssen, Werkzeuge zur Erzeugung von Metadaten oder auch Werkzeuge zur Konvertierung zwischen verschiedenen Metadaten-Schemas (Heery u. a., 2000).

Seit 1996 wird unter dem Telematics Application Programme der EU das DESIRE-Projekt¹⁴ gefördert, an dem Institute aus den Niederlanden, Norwegen, Schweden und Großbritannien beteiligt sind. Ziel dieses Projektes ist die Entwicklung und Explo-

¹⁴<http://www.desire.org>, Zugriffsdatum: 07.09.2000

<i>CORC</i>	<i>Definition CORC</i>	<i>DC-Element</i>	<i>DC Definition</i>
Title	Main Title	DC.Title	A name given to the resource
TitleAlternative	Added title	DC.Title.Alternative	Any form of the title used as a substitute or alternative to the formal title of the resource.
...
CoverageGeo	Spatial	DC.Coverage.spatial	Spatial characteristics of the intellectual content of the resource.
CovrgGeocode	Geographic coordinates and codes/abbrev.		
...

Tabelle 4.2: Mapping zwischen CORC und Dublin Core (Ausschnitt), MetaForm (1997–2000)

ration von neuen Technologien zur Unterstützung der wissenschaftlichen Kommunikation über das Internet. Im Zuge von DESIRE wurde auch eine prototypische Metadata-Registry entwickelt, die allerdings bisher nicht maschinenlesbar ist, sondern nur über ein Web-Interface den Zugriff auf die abgelegten Schemas ermöglicht (Heery u. a., 2000).

Neben der reinen Funktionalität als Referenzquelle für Syntax und Semantik von Metadaten-Schemas, stellen Metadata Registries auch sogenannte *mappings* zwischen verschiedenen Metadaten-Vokabularien zur Verfügung. Ein *mapping* zwischen zwei Vokabularien bedeutet dabei, dass zu jedem Element aus Vokabular A die Beziehung zu einem Element aus Vokabular B hergestellt wird. Tabelle 4.2.3.2 zeigt ein Mapping¹⁵ zwischen den Metadaten-Elementen von CORC (Cooperative Online Resource Catalog)¹⁶ und Dublin Core.

Im DESIRE-Projekt wird kein 1:1-Mapping zwischen jeweils zwei Metadaten-Schemas abgelegt, sondern jedes Vokabular wird auf eine *semantic layer* abgebildet (vgl. Abbildung 4.5). Der Vorteil dieses Verfahrens ist, dass nicht zwischen allen Schemas eine Mapping-Tabelle aufgebaut werden muss, sondern nur zwischen einem Schema und der *semantic layer*. Heery u. a. (2000) weist allerdings darauf hin, dass eine

¹⁵<http://dbl-www.sub.uni-goettingen.de/servlets/metaformList1Table=CORCDC&Head=CORC>, Zugriffsdatum: 12.09.2000

¹⁶CORC ist ein Service, der von der OCLC (Online Computer Library Center) in Zusammenarbeit mit mehreren Bibliotheken entwickelt wurde, um Bibliotheken Werkzeuge zur Verfügung zu stellen, um ihren Nutzern einen geführten Zugang zum Internet zu ermöglichen (<http://www.oclc.org/oclc/corc/faq/faq.htm>, Zugriffsdatum: 30.10.2000)

komplexe und detaillierte *semantic layer* notwendig ist, um eine Qualität zu erreichen, die manuell erstellten Mappings ebenbürtig ist.

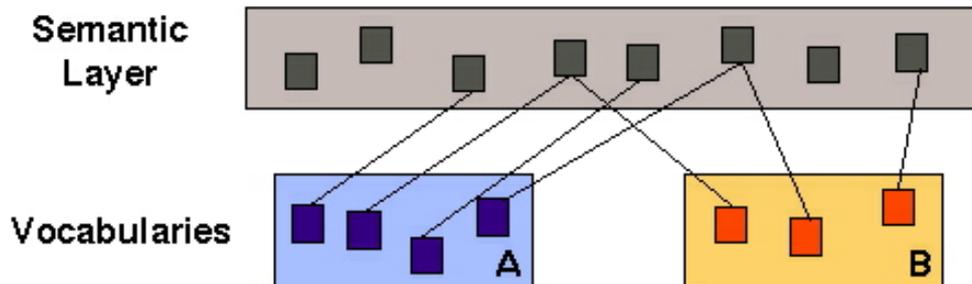


Abbildung 4.5: Abbildung von zwei Vokabularien aufeinander über eine *Semantic Layer* (Heery u. a., 2000)

4.2.3.3 Modulare Metadaten: Das Warwick Framework

Ein anderer Ansatz zur Erreichung von Interoperabilität ist das *Warwick Framework*, ein Ergebnis des zweiten Dublin Core Workshops an der Universität von Warwick 1996 (Dempsey und Weibel, 1996). Das Ziel besteht darin, unterschiedliche Arten von Metadaten, wie deskriptive Metadaten, Metadaten für das Zugangsmanagement etc., die zu einem digitalen Objekt gehören, in sogenannten *packages* zu organisieren, z.B. ein *package* für die Dublin Core Metadaten und eines für Terms & Conditions Metadaten. Der Vorteil dieser Methode gegenüber der Speicherung von einem Metadaten-Set, das alle Anwendungsfälle abdeckt, liegt in der Interoperabilität: So kann eine Anwendung, z.B. ein Web-Robot, die vorhandenen Dublin Core Metadaten eines Objekts auswerten, obwohl sie mit den speziellen Terms & Conditions Metadaten eines Objekts nichts anfangen kann. Die Architektur des Warwick Framework und beispielhafte Implementierungen in HTML, MIME und SGML werden ausführlich in Lagoze u. a. (1996) beschrieben. Das in 4.2.4 beschriebene Resource Description Framework eröffnet die Möglichkeit der Implementierung der Prinzipien des Warwick Framework (Causton, 1998, S. 36).

4.2.4 Resource Description Framework

Das *Resource Description Framework (RDF)* bietet ein domänen-neutrales Datenmodell, das die Kodierung, den Austausch und die Wiederverwendung von strukturierten Metadaten unterstützt. Gegenwärtig existiert zu RDF eine *Recommendation* des W3-Consortiums (REC-rdf-syntax, 1999), die das Datenmodell und eine mögliche Syntax für die maschinenlesbare Kodierung der Metadaten bietet und eine *Candidate Recom-*

mentation (rdf-schema, 2000), die spezifiziert, wie verschiedene Vokabularien in RDF zu formulieren sind .

RDF ist ein wichtiger Baustein für das sogenannte *Semantic Web*, das eine automatische Verarbeitung von Informationen im World Wide Web ermöglichen soll (Berners-Lee, 1998).

Lassila (1998) sieht Anwendungsmöglichkeiten für RDF in folgenden Bereichen:

- Unterstützung des Resource Discovery
- Katalogisierung und Beschreibung von Inhalten und der inhaltlichen Beziehungen für eine Website (automatische Generierung von Sitemaps) oder Dokumenten, die aus mehreren Objekten zusammengesetzt sind
- Benutzung durch Agentensysteme, die eigenständig Wissen austauschen
- Anwendung im *content rating* für den Schutz von Kindern vor unerwünschten Inhalten und zum Schutz der Privatsphäre von Internetnutzern
- Beschreibung von Urheberrechten an Web-Seiten
- Digitale Signaturen

4.2.4.1 Das Datenmodell des Resource Description Framework

RDF besteht aus *Knoten* und zugeordneten *Attribut-Wert*-Paaren. Ein *Knoten* kann jede beliebige Web-Ressource sein (identifiziert durch einen *Uniform Resource Identifier*¹⁷), Attribute sind der Ressource zugeordnete Eigenschaften und Werte sind entweder atomisch (i.e. Text-String) oder andere *Knoten* . Ein Tripel bestehend aus *Knoten*, *Attribut* und *Wert* wird als *property* bezeichnet. Abbildung 4.6 zeigt die Darstellung eines Tripels als gerichteten Graphen.

Abbildung 4.7 zeigt als Wert für das Attribut „Autor“ eine Instanz mit den *properties* Email und Name.

Eine mächtige Eigenschaft von RDF ist die Möglichkeit, Aussagen über *properties* zu treffen. Dabei wird ein Tripel, eine *property*, in einen *Knoten* transformiert. Dieser Prozess heißt *reification*. Eine mögliche Anwendung für diese Eigenschaft ist das Peer Review: Ein Gutachter trifft eine Aussage über die Qualität einer Web-Ressource. Bei einer solchen Anwendung ist allerdings die Einbeziehung von digitalen Signaturen

¹⁷Unter der Bezeichnung *Uniform Resource Identifier (URI)* werden von der *Internet Engineering Task Force (IETF)* die zukünftigen Bausteine der Internet-Architektur gefasst, die bekannte URL und die in 4.3.4 beschriebenen Uniform Resource Names (URN).

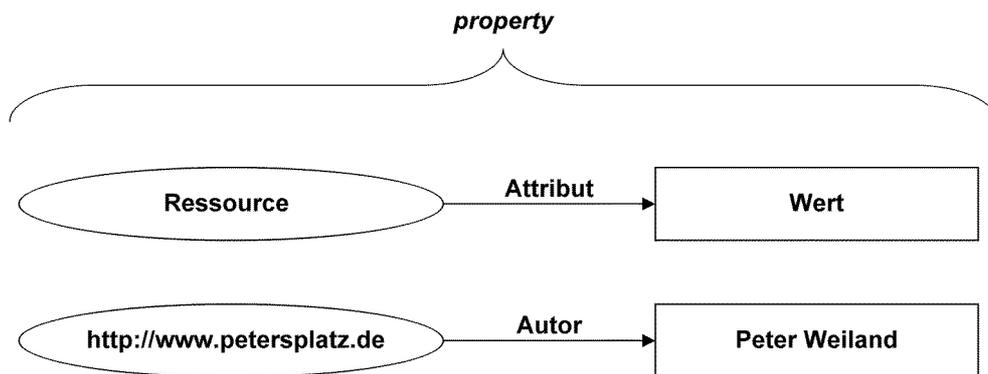


Abbildung 4.6: Graphendarstellung eines einfachen Tripels

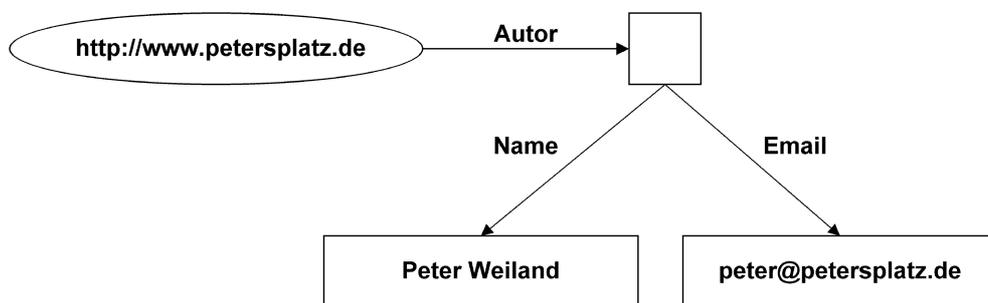


Abbildung 4.7: Graphendarstellung eines Tripels mit komplexem Wert für ein Attribut

sinnvoll. Abbildung 4.8 erweitert die in Abbildung 4.6 getroffene Aussage dahingehend, dass gesagt wird, von wem die Aussage stammt.

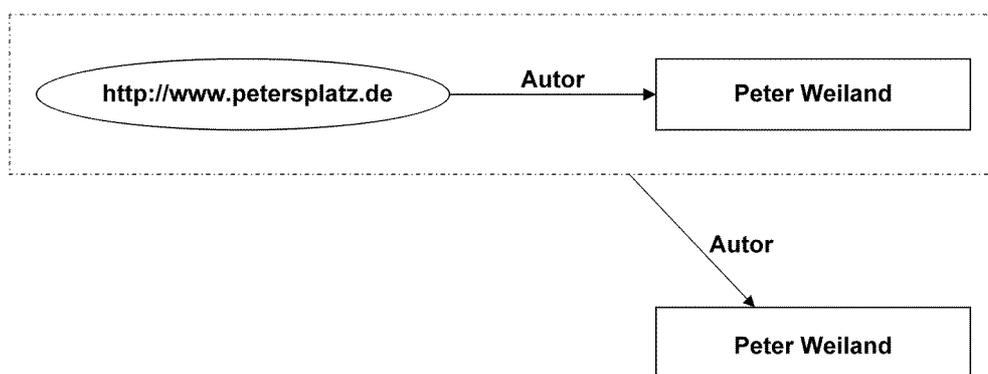


Abbildung 4.8: Aussage über eine Aussage

4.2.4.2 Die Syntax des Resource Description Framework

Zur Umsetzung des oben beschriebenen Datenmodells in eine maschinenlesbare Form ist eine Syntax notwendig. Die *eXtensible Markup Language (XML)* wird zur Repräsentation des beschriebenen Datenmodells benutzt.

RDF ermöglicht die Definition von Semantik für verschiedenste Anwendungen in der Beschreibung von Ressourcen. Da in verschiedenen Anwendungen gleiche Attribute unterschiedliche Bedeutungen haben können, benutzt RDF den *XML Namespace Mechanismus* zur eindeutigen Identifikation eines Vokabulars. Ein mögliches Vokabular für die Einbindung in RDF-Metadaten über den Namespace Mechanismus ist Dublin Core.

Folgendes Beispiel zeigt die Umsetzung des Graphen in Abbildung 4.6 in eine XML-Repräsentation.

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF" prefix="RDF"?>
<?xml:namespace ns = "http://purl.org/DC" prefix="DC"?>
```

Hier werden die Schemas RDF und Dublin Core in eine Beschreibung eingebunden. Die Einbindung von RDF ist notwendig, damit das RDF-Datenmodell in XML dargestellt werden kann, die Einbindung des Dublin Core Schema erlaubt die Verwendung dieses Vokabulars in der folgenden Beschreibung.

```
<RDF:RDF>
  <RDF:Description RDF:HREF = "http://www.petersplatz.de">
    <DC:Creator>Peter Weiland</DC:Creator>
  </RDF:Description>
</RDF:RDF>
```

Die Elemente `<RDF:RDF>` und `</RDF:RDF>` markieren die Grenzen der Beschreibung, die sich auf das RDF-Datenmodell bezieht. `<RDF:Description>` bezeichnet die Ressource, auf die sich die Beschreibung bezieht. Das Element `<DC:Creator>` repräsentiert das Dublin Core Attribut CREATOR mit dem Wert Peter Weiland.

Die Stärke von RDF liegt in der Zusammenarbeit unterschiedlicher Metadaten-Schemas. Miller (1998) gibt ein Beispiel für die Vermischung von Dublin Core mit einem fiktiven „Business Card“-Schema. Das Dublin Core Attribut CREATOR zeigt auf eine Beschreibung im „BusinessCard“-Schema.

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF/"
prefix = "RDF" ?>
<?xml:namespace ns = "http://purl.oclc.org/DC/"
prefix = "DC" ?>
<?xml:namespace ns = "http://person.org/BusinessCard/"
prefix = "CARD" ?>

<RDF:RDF>
  <RDF:Description RDF:HREF = "http://uri-of-Document-1">
    <DC:Creator RDF:HREF = "#Creator_001"/>
```

```
</RDF:Description>

<RDF:Description ID="Creator_001">
  <CARD:Name>John Smith</CARD:Name>
  <CARD:Email>smith@home.net</CARD:Email>
  <CARD:Affiliation>Home, Inc.</CARD:Affiliation>
</RDF:Description>
</RDF:RDF>
```

4.2.4.3 RDF Schema

Ein wichtiges Element innerhalb des Resource Description Framework ist die *RDF Schema Language* (rdf-schema, 2000), die dazu dient, für bestimmte Anwendungen Vokabularien zu definieren. Ziel ist die Formulierung verschiedenster Vokabularien, wie z.B. Dublin Core, in der RDF Schema Language, um die automatische Einbindung dieser Vokabularien in RDF-Beschreibungen mithilfe des Namespace-Mechanismus zu ermöglichen.

Die *Dublin Core Metadata Initiative* pflegt seit kurzer Zeit eine Registry mit verschiedenen Metadaten-Schemas in RDF¹⁸.

4.2.4.4 Werkzeuge für den Einsatz von RDF

Einen guten Überblick über die verfügbaren Werkzeuge zur Erstellung und Manipulation von RDF-Datenmodellen gibt die RDF-Webpage¹⁹ des W3C. Neben Parsern, die die XML-Repräsentation von RDF-Daten in eine abstrakte, auf dem RDF-Datenmodell basierende Form bringen, existieren auch einige Softwareprogramme zur Speicherung, Abfrage und Visualisierung von RDF-Metadaten.

Ein Projekt, das direkt auf den Einsatz von RDF im Kontext der Suche in RDF-Metadaten abzielt, ist das *EOR (Extensible Open RDF) Toolkit*²⁰.

4.2.5 Technische Infrastruktur

Der Erfolg eines Metadaten-Schemas wie Dublin Core ist in hohem Maße von der Verfügbarkeit leistungsfähiger Werkzeuge für die Erzeugung von Beschreibungen und zur Indexierung dieser Beschreibungen mit den dazugehörigen Informationsobjekten abhängig.

¹⁸<http://rdf.dev.oclc.org/myrdf/services/EOpenRegistry>, Zugriffsdatum: 13.11.2000

¹⁹<http://www.w3.org/RDF/>, Zugriffsdatum: 23.10.2000

²⁰<http://eor.dublincore.org/index.html>, Zugriffsdatum: 23.10.2000

4.2.5.1 Suchmaschinen

Laut der DC-FAQ (2000) unterstützen gegenwärtig folgende Suchmaschinen die Indizierung von in HTML-kodierten Dublin Core Metadaten:

- Ultraseek
- Swish-E
- Microsoft's Index Server
- Autonomy Knowledge Server
- Blue Angel Technologies MetaStar
- Verity Search 97 Information Server

Besonders geeignet für die Suche in verteilten Datenbeständen ist die *Harvest*-Architektur²¹, die in modifizierter Form auch im MathNet eingesetzt wird (Math-Net, 1999, S. 45). Harvest besteht aus den Komponenten *Gatherer* und *Broker*. Gatherer indexieren Web-Server; dabei können sie entweder auf demselben Rechner wie das Angebot oder auf einem entfernten Rechner installiert sein. Die Daten werden in ein Standardformat umgewandelt und einem Broker übermittelt. Der Broker sammelt die Indexe von einem oder mehreren Gatherers und baut einen kombinierten Index auf. Harvest ist hierarchisch aufgebaut, d.h. die kombinierten Indexe von mehreren Brokern können von einem anderen Broker wiederum zusammengefasst werden und dann über einen User Agent (z.B. Web-Browser) durchsucht werden. Die Vorteile von Harvest liegen zum einen in der geringeren Netzlast im Vergleich zu Web-Crawlers, und zum anderen in einer besseren Unterstützung des Information Retrieval, da die lokal installierten Gatherer-Komponenten individuell auf die jeweiligen Angebote hin konfiguriert werden können. Die Architektur von Harvest wird detailliert in Borggraefe (1999) beschrieben.

Weibel (2000a) führte kürzlich in einer Email an die Liste [dc-general] Gründe für die dürftige Unterstützung der Suchdienst-Anbieter für Dublin Core auf: Alle Suchdienst-Anbieter haben Verträge mit verschiedenen Firmen, die in den Suchergebnissen insbesondere ihre Sites vertreten sehen möchten, wobei diese Geschäftspolitik mal mehr und mal weniger publik gemacht wird. Desweiteren sehen viele Suchdienst-Anbieter die Sammlung von Metadaten als eher nutzlos an, da die Qualität der Metadaten keinerlei Kontrolle unterliegt.

²¹<http://www.tardis.ac.uk/harvest/>, Zugriffsdatum: 11.11.2000

Eine Alternative zum eigenständigen Sammeln von Metadaten durch Suchdienst-Anbieter ist die Integration von kontrollierten Metadaten-Kollektionen in die Datenbanken von Suchdienst-Anbietern. Hier stellt sich allerdings die Frage, inwieweit die Betreiber von Metadaten-Kollektionen bereit sind, diese öffentlich zugänglich zu machen.

4.2.5.2 Metadaten-Editoren

Im Gegensatz zu den Suchmaschinen sieht die Unterstützung für Dublin Core Metadaten auf der Editorenseite besser aus. Auf der Website der Dublin Core Metadata Initiative wird eine Aufstellung über frei verfügbare und kommerzielle Tools geboten²².

4.2.6 Diskussion

In der Vergangenheit wurden einige Kritikpunkte an Dublin Core geäußert. In Martin (1998) wurde unter anderem das Fehlen von formellen Standardisierungsbemühungen, die unzureichende Standardisierung der Qualifiers, das Fehlen von Implementierungsrichtlinien und die unzureichende Unterstützung durch Suchmaschinenbetreiber bemängelt. Diese für den Stand von 1997 geltenden Kritikpunkte sind heute weitgehend ausgeräumt (Weibel, 2000b); die Gründe für die mangelnde Unterstützung der Suchmaschinenbetreiber wurden oben bereits dargelegt. Martin (1998) kritisiert darüber hinaus die mangelnde Einbindung des Verlagssektors in die Entwicklung.

Der Erfolg von Dublin Core als Metadaten-Schema lässt sich an der Vielzahl der Projekte aus unterschiedlichen Disziplinen und unterschiedlichen Ländern ablesen²³. Eine interessante Beobachtung stellt in diesem Zusammenhang Caplan (1997) auf: Zu Beginn der Entwicklung lag der Schwerpunkt auf dem Einsatz von Dublin Core durch Autoren, mittlerweile wird jedoch die Mehrzahl an DC-Metadaten von Spezialisten auf dem Gebiet der Katalogisierung produziert.

²²<http://purl.org/DC/tools/index.htm>, Zugriffsdatum: 13.10.2000

²³<http://purl.org/DC/projects/index.htm>, Zugriffsdatum: 14.11.2000

4.3 Identifikation und Lokation von Web-Ressourcen

Die Notwendigkeit des Einsatzes von Identifiers zur eindeutigen Kennzeichnung und des Retrieval von Informationsobjekten wurde bereits in 3.3.2 und 3.3.3 begründet. Im folgenden werden die notwendigen Charakteristika von Identifiers für Web-Ressourcen beschrieben, und mit der URN und dem DOI werden zwei Implementierungen vorgestellt.

4.3.1 Reichweite von Identifiers

Ein wichtiger Aspekt bei der Diskussion um Identifiers ist die Frage, was ein Identifier überhaupt identifiziert. Miller (2000a) illustriert dieses Problem am Beispiel des Buches „Der Name der Rose“ von Umberto Eco. Ausgehend von der Definition der *International Federation of Library Associations (IFLA)* existiert „Der Name der Rose“ als *work*, i.e. *a distinct intellectual or artistic creation* (FRBR, 1998, S. 16). Die nächste Hierarchiestufe ist die sogenannte *expression*. Eine *expression* wird definiert als *the intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms* (FRBR, 1998, S. 18). Eine *expression* bezeichnet nicht das Werk als ein abstraktes Konzept, sondern die tatsächliche Realisierung z.B. in Form eines niedergeschriebenen Textes oder auch der Übersetzung eines Textes. Bezogen auf das Werk „Der Name der Rose“ bedeutet *expression* der Text des Werkes. Eine weitere *expression* ist z.B. eine Übersetzung in eine andere Sprache, allerdings bleibt das Werk das gleiche. Unterhalb der *expression* existiert die *manifestation*, die definiert ist als *the physical embodiment of an expression of a work* (FRBR, 1998, S. 20). Für „Der Name der Rose“ ist *manifestation* die konkrete Umsetzung der *expression* in eine bestimmte Ausgabe. Eine Suche bei dem Buchversender Amazon liefert vier ISBN-Nummern für „Der Name der Rose“, d.h. es existieren in der Datenbank von Amazon vier verschiedene *manifestations* dieses Werkes.

Dieses Beispiel zeigt, dass die ISBN kein Identifier für einen bestimmten Inhalt darstellt, sondern nur *manifestations*, i.e. bestimmte Ausgaben identifizieren kann. Für einen an einem bestimmten Werk Interessierten ist die *manifestation* oft zweitrangig: Sie/Er ist an dem Werk (oder auch an der *expression*) interessiert. Tatsächlich in Händen halten kann man dann allerdings nur ein *item*, i.e. *a single exemplar of a manifestation* (FRBR, 1998, S. 23).

Das Problem, was ein Identifier tatsächlich identifiziert, stellt sich in besonderer Weise auch für die elektronische Kommunikation: Von einem elektronischen Aufsatz

können mehrere *expressions* existieren, z.B. eine Langversion und eine gekürzte Version oder verschiedene Revisionen des gleichen Aufsatzes. Darüber hinaus können diese *expressions* wiederum in verschiedenen *manifestations* vorliegen: z.B. im PDF-Format, im PostScript-Format oder auch als HTML. Die nächsttiefere Hierarchieebene im Referenzmodell der IFLA, das *item*, kann hier ebenfalls auftreten als an verschiedenen Orten vorgehaltene elektronische Kopien des Aufsatzes.

Ein weiteres Problem, dass sich bei Verwendung der ISBN (oder auch in noch stärkerem Maße bei der ISSN) stellt, ist die *Granularität* der identifizierten Information. Die ISBN ist beschränkt auf die Buchebene, mit der ISSN können sogar nur Zeitschriftentitel identifiziert werden.

4.3.2 Warum die URL kein Identifier ist

Mit dem Aufkommen des World Wide Web haben sich die *Uniform Resource Locators (URL)* als Technologie zum Retrieval von Inhalten etabliert. Wie oben beschrieben, bieten URLs keine Spezifizierung eines Inhaltes (z.B. eines bestimmten Aufsatzes), sondern sind nur Anweisungen, wie auf ein bestimmtes Objekt zugegriffen werden kann. URLs bestehen aus dem Namen des Dienstes, der angesprochen werden soll (z.B. *ftp* für *File Transfer Protocol* oder *http* für das *hypertext transfer protocol*, das zum Zugang zu Web-Seiten dient), dem Namen des Servers und einer Datei auf diesem Server (z.B. `http://is.uni-sb.de/index.html` für die Datei `index.html` auf dem Server mit dem Namen `is.uni-sb.de`). Da die Datei `index.html` ständig Änderungen unterworfen ist, kann mit der angegebenen URL kein bestimmter Inhalt identifiziert werden. Die URL ist somit ein Locator und kein Identifier.

Für die bibliographische Kontrolle und das Resource Discovery ist der Einsatz dauerhaft gültiger Bezeichnungen für Inhalte notwendig, die URL als Locator von Inhalten ist hier nicht ausreichend.

4.3.3 Anforderungen an Identifiers

In den beiden vorangegangenen Abschnitten wurde aufgezeigt, aus welchen Gründen die sogenannten *legacy identifiers* (Bide, 1997, S. 3) wie die ISBN und ISSN und die URL für die Identifikation in der wissenschaftlichen Kommunikation nicht ausreichend sind. Sollins und Masinter (1994) stellen für die Uniform Resource Names (URN) acht Anforderungen an Identifiers:

Global scope A URN is a name with global scope which does not imply a location. It

has the same meaning everywhere.

Global uniqueness The same URN will never be assigned to two different resources.

Persistence It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.

Scalability URNs can be assigned to any resource that might conceivably be available on the network, for hundreds of years.

Legacy support The scheme must permit the support of existing legacy naming systems, insofar as they satisfy the other requirements described here. For example, ISBN numbers, ISO public identifiers, and UPC product codes seem to satisfy the functional requirements, and allow an embedding that satisfies the syntactic requirements described here.

Extensibility Any scheme for URNs must permit future extensions to the scheme.

Independence It is solely the responsibility of a name issuing authority to determine the conditions under which it will issue a name.

Resolution A URN will not impede resolution (translation into a URL, q.v.). To be more specific, for URNs that have corresponding URLs, there must be some feasible mechanism to translate a URN to a URL.

Diese Anforderungen gelten auch für andere Identifiers wie den DOI (vgl. Paskin, 1999).

4.3.4 Uniform Resource Names

4.3.4.1 Definition

Schon 1994 wurde die Notwendigkeit für die Identifizierung (im Gegensatz zur Lokalisierung durch URLs) von Ressourcen im Internet in einem *Request for comment* festgehalten. Sollins und Masinter (1994) definieren die Reichweite der vorgeschlagenen Uniform Resource Names (URN) und grenzen die URNs von den URLs ab:

A URN identifies a resource or unit of information. It may identify, for example, intellectual content, a particular presentation of intellectual content, or whatever a name assignment authority determines is a distinctly namable entity. A URL identifies the location or a container for an

instance of a resource identified by a URN. The resource identified by a URN may reside in one or more locations at any given time, may move, or may not be available at all. Of course, not all resources will move during their lifetimes, and not all resources, although identifiable and identified by a URN will be instantiated at any given time. As such a URL is identifying a place where a resource may reside, or a container, as distinct from the resource itself identified by the URN.

4.3.4.2 Framework für die Anwendung von URNs

Aus der obigen Definition geht hervor, dass eine URN kein Locator wie die URL ist, d.h. aus der URN geht der Speicherort eines Objektes nicht hervor. Daher ist ein über das Internet zugänglicher Dienst notwendig, der eine URN mit einer korrespondierenden Ressource in Verbindung bringt. Dieser Prozess wird *Resolution* genannt. Das Ergebnis des Resolution-Prozesses ist gewöhnlich eine URL oder eine Liste mit URLs, die den Speicherort des Objektes angeben.

Arms u. a. (1996) beschreiben ein Framework für die Anwendung von URNs auf einer breiten Basis. Unterschieden wird zwischen *naming schemes* und *resolution systems*. *Naming schemes* sind Prozeduren für die Schaffung und Vergabe eindeutiger URNs, die einer definierten Syntax unterliegen; *resolution systems* dienen der oben beschriebenen Auflösung von URNs. Betont wird die Unabhängigkeit zwischen *naming schemes* und *resolution systems*: Ein *naming scheme* ist nicht gebunden an ein spezifisches *resolution system*. Bedingt durch diese Unabhängigkeit ist die Existenz von *URN registries* notwendig, mit denen ein Nutzer diejenigen *resolution systems* finden kann, die eine bestimmte URN auflösen können. Das Konzept der URN registries wird von Sollins (1998) als *Resolver Discovery Service (RDS)* etabliert.

4.3.4.3 Syntax der URN

Die Syntax der URN wurde in Moats (1997) definiert. Eine URN hat danach die Form:

$$\langle \text{URN} \rangle ::= \text{"urn:"} \langle \text{NID} \rangle \text{" : " } \langle \text{NSS} \rangle$$

Tabelle 4.3: Syntax der URN

"urn:" ist dabei ein allgemeiner Bezeichner, um eine URN als solche auch innerhalb eines Textes für einen Menschen oder einen Parser kenntlich zu machen; <NID> ist der *Namespace Identifier* und <NSS> ist der *Namespace Specific String*, i.e. der Name einer Ressource innerhalb des angegebenen Namespace. Die Anwendung der

URN auf bekannte Namespaces wie ISBN verdeutlicht den Aufbau (zur Benutzung existierender bibliographischer Identifiers als URNs siehe Lynch u. a. 1998):

```
urn:isbn:0262011808
```

Tabelle 4.4: Fiktives Beispiel der Kodierung einer ISBN in eine URN

Die Namespaces sind ein wichtiges Konzept innerhalb des URN Frameworks. Die Beanspruchung eines bestimmten *namespace identifier* durch ein *naming scheme* muss einhergehen mit der Registrierung dieses *namespace identifier* durch die *Internet Assigned Numbers Authority (IANA)*²⁴. Damit ist sichergestellt, dass eine bestimmte Organisation (mit einem naming scheme) einen bestimmten Namespace für sich beanspruchen kann.

Innerhalb eines Namespaces ist die Vergabe der *namespace specific strings* freigestellt. Es muss allerdings sichergestellt sein, dass eine URN eine Ressource eindeutig definiert.

4.3.4.4 URN-Resolution

Die Auflösung von URNs zu den korrespondierenden Ressourcen geschieht über ein netzwerk-zugängliches *resolution system*. Da nicht jeder Resolver jede URN auflösen kann, ist das *Resolver Discovery System (RDS)* notwendig. Mithilfe des RDS lässt sich für eine URN ein geeignetes *resolution system* finden, d.h. der Client, der eine URN an ein RDS sendet, wird an ein geeignetes *resolution system* verwiesen, bzw. wird direkt auf die Ressource geleitet. Dieses globale Modell der Auflösung von URNs zu URLs stellt neben den technischen Herausforderungen, hohe Anforderungen hinsichtlich des Managements und der Registrierung von Namespaces und Resolver-Diensten. Die Prinzipien einer derartigen Architektur sind in Sollins (1998) beschrieben, die Grundlagen für die experimentelle Implementierung eines Resolver Discovery Systems mit Hilfe des Domain Name System (DNS) finden sich in Daniel und Mealling (1997).

Neben der globalen Auflösung von URNs (siehe Abb. 4.9) ist der Einsatz auch lokal, d.h. beschränkt auf eine bestimmte Organisation mit ihren Mitgliedern, möglich. Technisch lässt sich die lokale Auflösung mit Hilfe von Proxies realisieren: Eine URN, die ein User Agent (z.B. Web-Browser) absetzt und die einen innerhalb der Organisation verwendeten Namespace Identifier (NID) hat, wird von einem Proxy abgefangen und an einen lokalen Resolver weitergeleitet. Der lokale Resolver leitet die Anfrage

²⁴<http://www.iana.org>, Zugriffsdatum: 15.09.2000

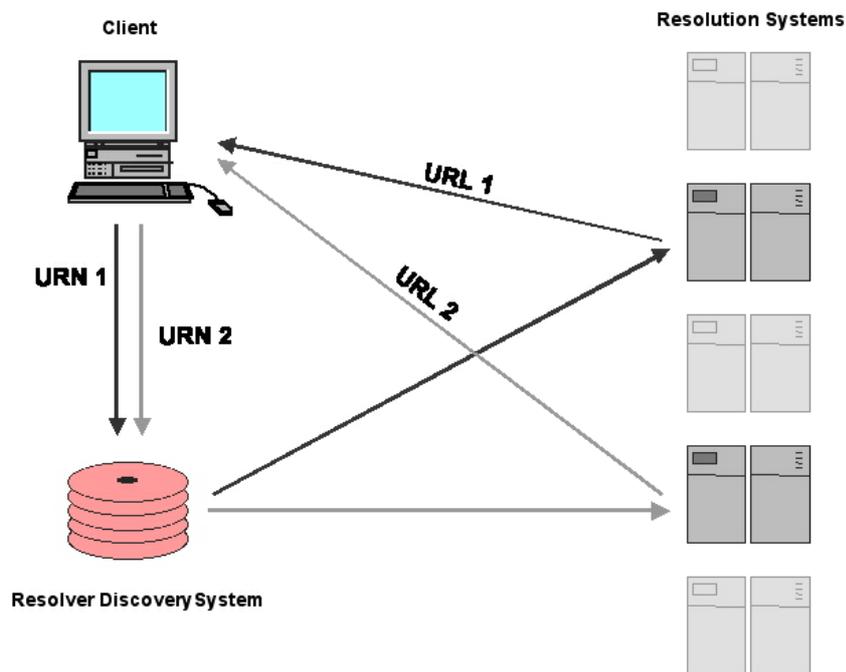


Abbildung 4.9: Globale Auflösung von URNs mithilfe des Resolver Discovery System

des User Agent an den Speicherort (z.B. URL) der gewünschten Ressource (van der Werf-Davelaar, 1999).

4.3.4.5 Der URN-Generator des Nordic Metadata Projects

Im Rahmen des Nordic Metadata Projects werden elektronische Quellen mit Hilfe von URNs identifiziert. Die Bibliothek der Universität von Helsinki ließ dafür die *National Bibliographic Number (NBN)* als Namespace Identifier bei der IANA reservieren. Die NBN wird weltweit von Nationalbibliotheken vergeben, wobei keine standardisierten Vergaberichtlinien existieren. Die Nationalbibliotheken von Finnland, Norwegen und Schweden bieten web-basierte URN-Generatoren, mit denen Autoren und Herausgeber eindeutige Identifiers für ihre elektronischen Dokumente erhalten können. Diese Dokumente werden von den jeweiligen Bibliotheken archiviert. Ein Auflösungsmechanismus für diese URNs existiert bisher nicht, allerdings werden die NBNs über den Nordic Web Index verfügbar gemacht (siehe Hakala u. a., 1998, S.29).

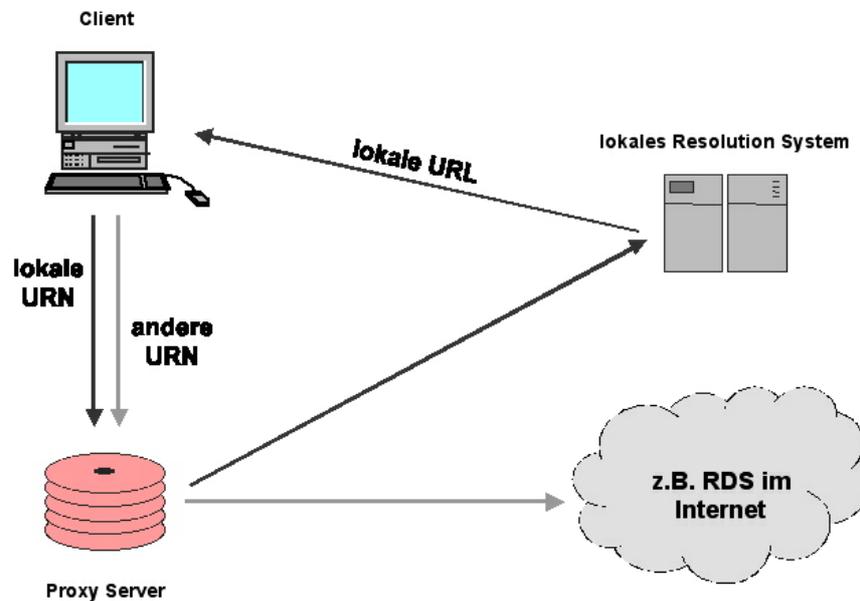


Abbildung 4.10: Lokale Auflösung von URNs mithilfe eines Proxy-Servers

4.3.5 Digital Object Identifier (DOI)

4.3.5.1 Ursprung und Entwicklung

Der Digital Object Identifier (DOI) geht auf Bestrebungen der *Association of American Publishers (AAP)*²⁵ zurück, ein System zu etablieren, das einerseits dem Urheberrechtsschutz dient, andererseits geschäftliche Transaktionen ermöglicht. Rosenblatt (1997) nennt die sechs Anforderungen, die sich an ein solches System stellen:

1. *Content identification*, a "tag" on each item that identifies it uniquely and over time.
2. *Content description*, such as bibliographic data and key words, that enables searching and browsing.
3. *Access rights* that state, for example, whether a user may view, print, or copy an item, whether he or she is buying, borrowing, or accepting transference of rights, and under what conditions those rights are conveyed.

²⁵<http://www.publishers.org>, Zugriffsdatum: 17.09.2000

4. *Display formats*, such as HTML, PDF, or XML for page-like content, or other formats for sound and video.
5. *Content protection schemes*, such as encryption and watermarking.
6. *Financial-transaction information*, such as account numbers, prices, etc.

Die Bemühungen um die Entwicklung eines Systems, das den aufgeführten Anforderungen genügt, mündete 1997 in der Gründung der International DOI Foundation (IDF), wobei der Fokus zuerst auf der Entwicklung eines *Content Identifiers* lag. Der Entwicklungsprozess wird von Paskin (2000, Seite 8-1) als *Three-Track-Approach* beschrieben. Der erste Schritt bestand in der Entwicklung eines Systems, das einen DOI zu einer URL auflöst. Der erste Demonstrator wurde im Oktober 1997 auf der Frankfurter Buchmesse vorgestellt, seitdem wurden über 3 Millionen DOIs vergeben. Die bisher vergebenen DOIs sind allerdings auf die sogenannte *single-resolution*, i.e. die Auflösung eines DOI zu genau einer URL, beschränkt. Der zweite Entwicklungsschritt ist die Auflösung eines DOI zu mehreren URLs (*multiple-resolution*) und die Verlinkung von Metadaten mit dem DOI, was eine Voraussetzung für das Management urheberrechtlich geschützter Quellen ist. Parallel zu beiden Entwicklungsschritten versucht die IDF mit anderen Organisationen zu kooperieren, die auf benachbarten Feldern arbeiten.

4.3.5.2 Was identifiziert ein DOI?

In 4.3.1 wurde mit dem Referenzmodell der IFLA eine Möglichkeit der Klassifizierung geistigen Eigentums vorgestellt. Das dem DOI zugrundeliegende Modell wurde im Rahmen des *indecs*-Projekts entwickelt und stellt eine abgewandelte Form des IFLA-Modells dar: *Item* wird im *indecs*-Modell als ein Typ der *manifestation* angesehen und *expression* ist eine Schöpfung, die in Raum und Zeit existiert, aber nicht physikalisch vorhanden ist (Bearman u. a., 1999). Abbildung 4.11 zeigt die Beziehungen zwischen den verschiedenen Typen von geistigem Eigentum.

Der DOI kann alle Entitäten des *indecs*-Modells identifizieren, da ein DOI auch zu einem anderen DOI aufgelöst werden kann (z.B. kann ein DOI für ein *work* zu mehreren DOIs aufgelöst werden, die *manifestations* von diesem *work* identifizieren).

In diesem Zusammenhang ist die sogenannte *Granularität* von Bedeutung: Prinzipiell ist die Größe oder die Art der Objekte, die ein DOI identifiziert, nicht festgelegt. Für die Publikation eines wissenschaftlichen Aufsatzes bedeutet dies, dass ein DOI

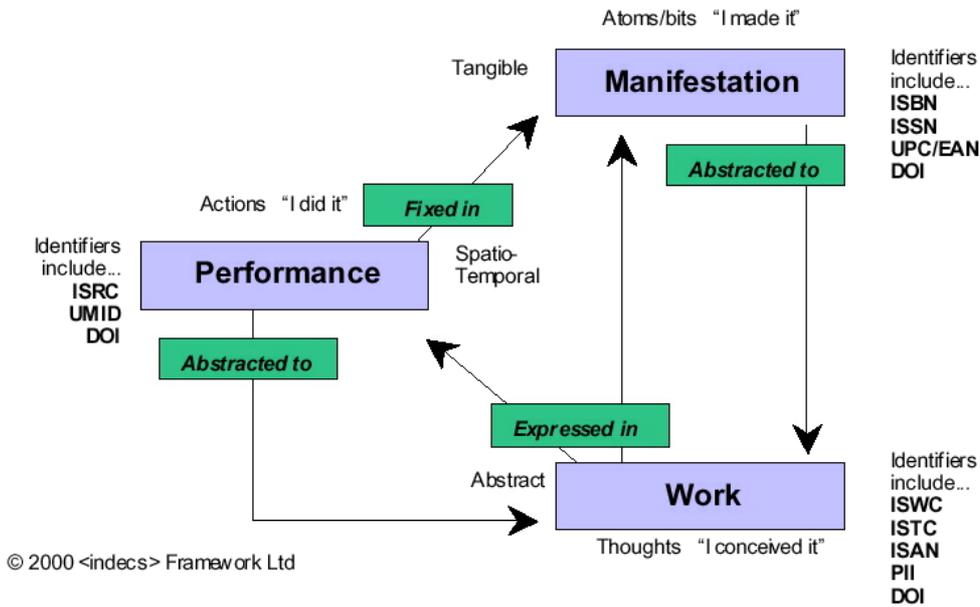


Abbildung 4.11: Beziehungen zwischen den verschiedenen Typen geistigen Eigentums (Rust und Bide, S. 23)

das gesamte Dokument identifiziert, genauso können auch einzelne Teile dieses Dokumentes mit einem DOI versehen werden, um beispielsweise eine bestimmte Abbildung zu identifizieren. Die oben erwähnte Eigenschaft der Resolution eines DOI zu mehreren anderen DOIs stellt sicher, dass der Zusammenhang des Dokumentes mit seinen Einzelbestandteilen erhalten bleibt.

4.3.5.3 Das DOI-System und seine Komponenten

Hinter dem DOI als Identifier, d.h. als Nummerierungsschema, steht ein komplexes System interagierender Komponenten. Das DOI-System besteht aus folgenden vier Komponenten:

- *Enumeration:* Vergabe eines Identifiers an ein geistiges Eigentum
- *Description:* Beschreibung einer Entität, die einen DOI besitzt
- *Resolution:* Datenbanken mit Abbildungen von DOIs auf andere DOIs oder URLs; Technologie zur Auflösung von DOIs
- *Policies:* Regeln, die die Prozessabläufe innerhalb des Systems definieren

4.3.5.4 Enumeration

Ein DOI besteht aus zwei Teilen: Einem *prefix*, das die Identifikation innerhalb des Handle-Systems und die Nummer der Organisation, die den DOI vergibt, enthält sowie einem *suffix*, das das Objekt eindeutig identifiziert. Tabelle 4.3.5.4 zeigt die Struktur eines DOI, wobei <DIR> den *Directory Code* bezeichnet, der für alle DOIs den Wert 10 hat, <REG> (*Registrant Code*) den Code der DOI-vergebenden Organisation bezeichnet und <DSS> (*DOI Suffix String*) für den von der Organisation vergebenen Bezeichner steht.

$$\langle \text{DIR} \rangle . \langle \text{REG} \rangle / \langle \text{DSS} \rangle$$

Tabelle 4.5: Syntax des DOI nach ANSI/NISO Z39.84-2000 (Paskin, 2000, Appendix 1.2)

Der DOI Suffix String kann von einer Organisation, die einen *Registrant Code* besitzt, frei vergeben werden. Diese Organisation ist dann für die Vergabe eindeutiger Bezeichner verantwortlich. In den DSS können auch andere Identifiers, wie z.B. eine ISBN integriert werden, allerdings ist nicht vorgesehen, den DSS interpretierbar zu machen, d.h. der DOI ist ein *unintelligenter* Identifier. Die Vergabe von unintelligenten Bezeichnern scheint auf den ersten Blick ein Nachteil zu sein. Bide und Green (1997) zeigen am Beispiel der ISBN die Nachteile **intelligenter** Identifier auf: Da in der ISBN der Verlag kodiert wird, liegt es nahe, die ISBN daraufhin auszuwerten. In der Praxis werden die Rechte an Büchern allerdings auch von Verlag zu Verlag übertragen, die ISBN des Buches bleibt dabei zunächst aber mal gleich. Diese Kritik ist generalisierbar auf andere Identifiers und es ist zweifelhaft, ob die in einen intelligenten Identifier kodierte Bedeutung langfristig Bestand haben kann.

4.3.5.5 Description

Der Fokus des DOI liegt nicht allein auf der persistenten Identifikation von Internet-Ressourcen, sondern auch auf erweiterten Anwendungen, die eine grundlegende Einbeziehung von Metadaten in den Identifier benötigen. Der DOI kann ein bestimmtes Objekt zwar eindeutig identifizieren, ohne Metadaten ist es aber nicht möglich festzustellen, was der Inhalt eines Objektes ist (es sei denn, man schaut sich das identifizierte Objekt direkt an). Mit Metadaten sind erweiterte Anwendungen wie elektronischer Handel von digitalen Objekten, die oben erwähnte *multiple resolution* in Verbindung mit einem intelligenten Selektionsprozess für die Quelle oder auch das in 4.4.1 beschriebene *reference linking* denkbar.

Die DOI-Metadaten basieren auf dem Datenmodell des indecs-Projektes. Dieses

Datenmodell bietet ein formale Struktur für die Beschreibung und eindeutige Identifizierung von geistigem Eigentum, wobei ein Schwerpunkt auf der Interoperabilität mit anderen Metadaten-Schemas, wie z.B. Dublin Core, liegt. Die Elemente des DOI Kernel Metadata Sets sind beschrieben in Paskin (2000, Seite 5-3). Neben der Registrierung der Metadaten bei der Vergabe eines DOI muss auch das Objekt mindestens einem *DOI Genre* zugeordnet werden. Ein Genre ist definiert als „*A class of creations, sharing common attributes, for which particular additions or qualifications to the DOI metadata kernel are mandated, either for general purposes or limited to specific functional requirements as designated by the Genre*“ (Paskin, 2000, Seite 5-4). Beispiele für Genre umfassen akademische Zeitschriftenartikel, Musikdateien im MP3-Format oder auch biomedizinische Fotografien. Genres können von jeder Organisation registriert werden, die das Recht zur Vergabe von DOIs hat.

4.3.5.6 Resolution

Resolution im Kontext der Identifiers wurde schon in 4.3.4.4 beschrieben. Die Auflösung von DOIs zu anderen DOIs oder URLs ist insofern einfacher, als dass es keine Notwendigkeit für den Einsatz eines Resolver Discovery Service gibt, da es innerhalb des Systems nur einen Namespace - nämlich DOI - gibt. Die einfache Auflösung eines DOI zu einer URL ist in Abbildung 4.12 dargestellt:

1. Der Benutzer stößt auf einen DOI und gibt diesen weiter an den DOI Directory Service.
2. Der DOI Directory Service liefert eine URL zurück.
3. Der Browser fordert die URL an.
4. Der unter der gesendeten URL liegende Inhalt wird an den Web-Browser zurückgesendet.

Wie oben erwähnt ist eine wesentliche Eigenschaft des DOI-Systems die Fähigkeit zur *multiple resolution*, d.h. ein DOI kann zu mehreren URLs, DOIs oder auch zu anderen Daten aufgelöst werden. Die Anwendungsfälle für *multiple resolution* sind vielfältig:

- Ein Dokument besteht aus mehreren Objekten, die alle durch einen DOI identifiziert werden.

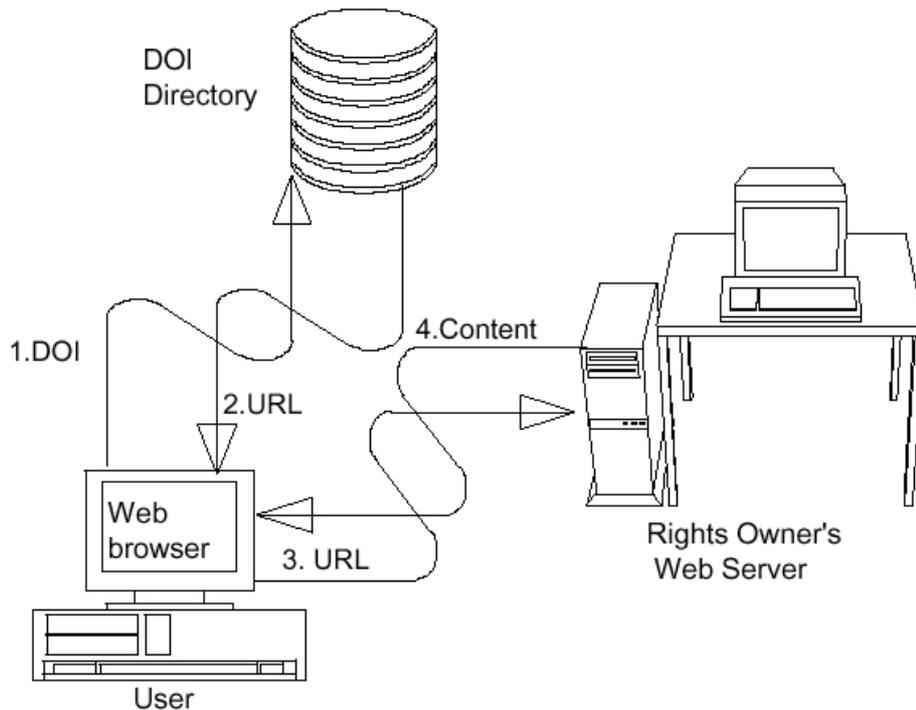


Abbildung 4.12: Auflösung eines DOI aus Sicht des Nutzers (Bide, 1997)

- Ein Dokument ist unter mehreren URLs abgelegt; der Auflösungsmechanismus sucht den Speicherort, der geographisch am nächsten zum Anforderungsort liegt.
- Der Zugriff auf ein bestimmtes Dokument ist kostenpflichtig: Ein nicht entsprechend autorisierter Benutzer kann den DOI nur zu den Metadaten oder dem Abstract des Dokumentes auflösen.

Die einfachste Form der *multiple resolution* ist die Auflösung des DOI zu einer Liste von URLs oder DOIs. Aus dieser Liste kann ein Benutzer dann die gewünschte Ressource auswählen. Die oben beschriebenen möglichen Anwendungsszenarien erfordern allerdings eine automatische Selektion des passenden Dienstes (siehe Abbildung 4.13).

Der DOI benutzt das *Handle System*, das schon 1994 als Grundlage für digitale Bibliothekssysteme von der *Corporation for National Research Initiatives (CNRI)* entwickelt wurde. Das Handle System ermöglicht die sichere Auflösung und Administration von persistenten Namen über das Internet. Die Elemente des Handle Systems sind dabei ein *namespace (hdl)*, ein zentral verwalteter Registrierungsservice (*Global*

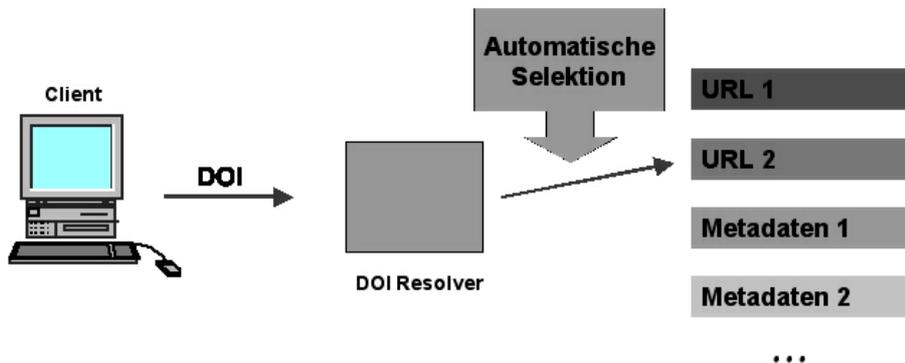


Abbildung 4.13: Multiple resolution eines DOI (Paskin, 2000, Seite 6-2)

Handle Registry) und Protokolle zur verteilten Speicherung und Auflösung der handles (Identifiers). Für den DOI ist im Handle-System der Namespace „10.“ reserviert. Die technischen Grundlagen des Handle-Systems werden in Sun und Lannom (2000) ausführlich beschrieben. Abbildung 4.14 illustriert den Ablauf der Auflösung einer Handle.

1. Ein Client (z.B. Web-Browser) stößt auf eine Handle (in der Abbildung 10.123/456) und sendet diese an das Handle-System, damit sie zu einer URL aufgelöst werden kann. Diese Weiterleitung kann entweder über einen Proxy-Server stattfinden, der eine Handle als solche erkennt und richtig weiterleitet oder über ein Browser-PlugIn, das das Handle-System direkt kontaktiert.
2. Das Handle-System besteht aus lokalen Handle-Services, die sich über mehrere Sites erstrecken können und einer globalen Handle Registry. Die globale Handle Registry verfügt über die Informationen, welche Namespaces von welchen Handle Services aufgelöst werden können. In diesem Schritt wird der geeignete lokale Handle-Service gefunden und die Anfrage an diesen weitergeleitet.
3. Die Handle 10.123/456 wird zu einer URL aufgelöst, wobei eine Handle auch zu mehreren URLs aufgelöst werden könnte.
4. Die URL wird an den Client zurückgesendet.

4.3.5.7 Policies

Das Funktionieren des DOI-Systems erfordert genau formulierte Regeln (Policies) für alle Beteiligten. Die Formulierung dieser Regeln ist die Aufgabe des Vorstandes der

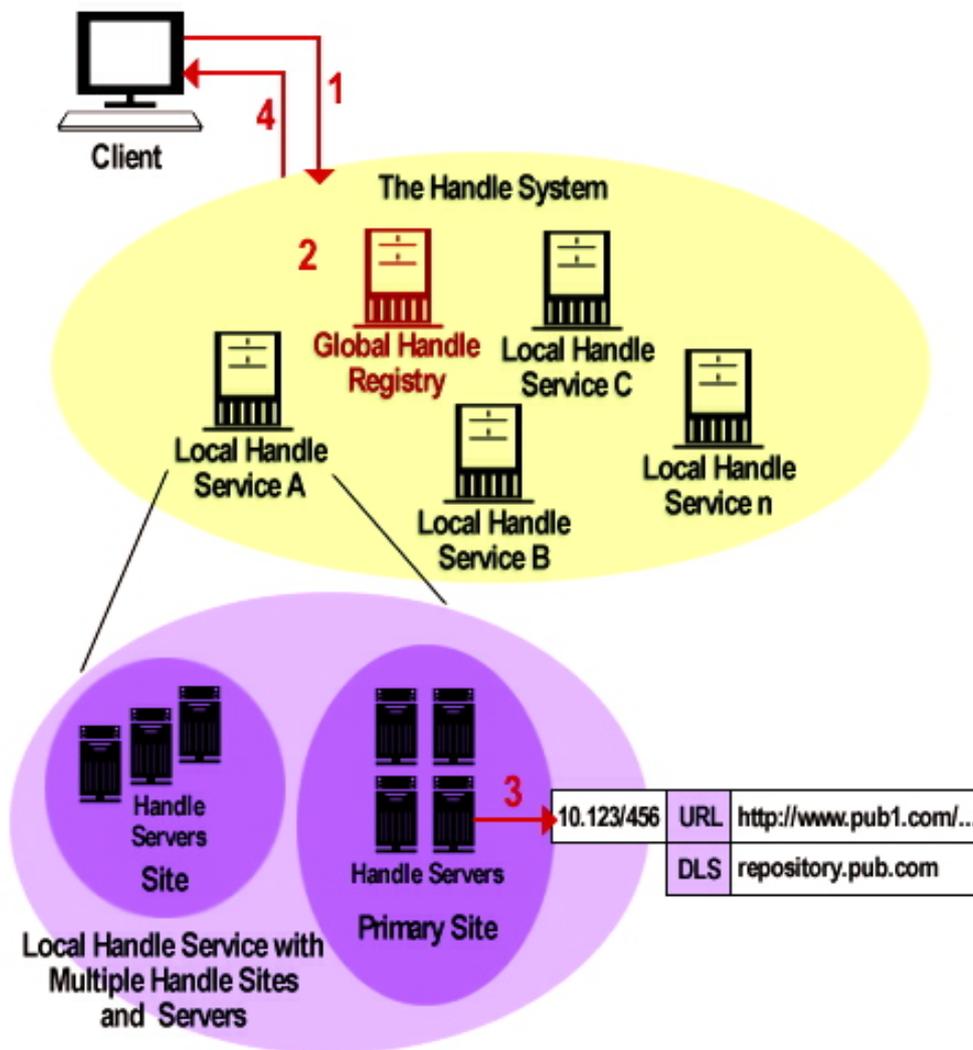


Abbildung 4.14: Architektur des Handle-Systems (Sun und Lannom, 2000)

International DOI Foundation. Der gegenwärtige Regelkatalog umfasst sechs Punkte (Paskin, 2000, Seite 7-1f.):

- The DOI can be used to identify any intellectual property entity. Our definition of intellectual property is a broad one, following that of WIPO agreements and similar international instruments; this includes both physical and digital manifestations, performances and abstract works. The primary focus of the DOI is on the management of intellectual property rights, but this does not preclude (for example) issuing a DOI to an entity that is in the public domain. An entity can be identified at any arbitrary level of granularity.

- All DOIs must be registered with the global DOI registry. Registrants are responsible for the maintenance of state data and metadata relating to DOIs that they have registered.
- The syntax of the DOI follows a standardised syntax. In use, the DOI is an opaque string (dumb number).
- Each entity registered for a DOI will be assigned by its Registrant to at least one DOI Genre; the rules relating to metadata declaration for that Genre will be followed. A minimum kernel of metadata will be declared for all DOIs registered. This kernel of metadata will be freely available to any user, so that a basic description of the entity that DOI identifies can be accessed. Other services dependent on DOI metadata will not necessarily be free; extensions to metadata beyond the kernel will not necessarily be publicly accessible.
- The use of the DOI system for resolution of a DOI should be free to users; the costs of operation of the system should be borne directly or indirectly by the Registrants.
- Registration Agencies will be established to manage the assignment of DOIs, their registration and the declaration of the metadata associated with them. The business model adopted by an individual Registration Agency is a matter for the Agency alone, so long as it complies with overall IDF policy.

4.3.5.8 Registrierung von DOIs

Gegenwärtig ist die International DOI Foundation die einzige Instanz, die DOI prefixes vergibt. Diese Registrierung²⁶ kostet 1000US\$ und beinhaltet den Zugang zu web-basierten Werkzeugen zur Generierung und Administration von DOIs. Für die Vergabe einer großen Anzahl von DOIs steht auch ein Stapelverarbeitungsprozess (Batch) zur Verfügung (Paskin, 2000). Der Gebrauch von DOIs ist gegenwärtig sowohl für Endnutzer als auch Autoren/Herausgeber kostenfrei.

Für die Zukunft ist die Schaffung von sogenannten *Registration Agencies* geplant, die für die Vergabe von DOI prefixes und die Registrierung von DOIs verantwortlich sind. Darüber hinaus stellen sie den Registrierenden die notwendige Infrastruktur für die Pflege von Metadaten und *state data* zur Verfügung. Im Gegensatz zu International DOI Foundation dürfen die Registration Agencies profitorientiert operie-

²⁶Die Registrierung kann über ein Web-Formular (<http://www.doi.org/started.html>, Zugriffsdatum: 22.09.2000) erfolgen.

ren, gegenwärtige Überlegungen für Geschäftsmodelle basieren auf der Erhebung von Gebühren für die Vergabe von DOIs, Gebühren für die Auflösung von DOIs oder auch indirekt durch kostenpflichtige Zusatzangebote (Paskin, 2000, Seite 11-1).

4.3.5.9 Verbreitung des DOI-Systems

Der International DOI Foundation gehören momentan 41 Organisationen an²⁷, darunter alle großen wissenschaftlichen Verlage.

4.3.6 Diskussion

Davidson und Douglas (1998) führen einige Kritikpunkte am DOI-System an, die vor allem auf die Funktionalität des DOI zur dauerhaften Identifizierung von Objekten im Internet zielen. Gegenwärtig wird das DOI-System in erster Linie von großen Verlagen für ihre Produkte genutzt, der beklagte Mangel an Persistenz von Links im Internet betrifft aber eher die Angebote von Einzelpersonen oder kleineren Organisationen. Daher ist der Effekt für die Stabilität von Links im Internet eher zu vernachlässigen. Ein weiterer Punkt betrifft die Kosten für die Pflege der DOIs und die mit dem Objekt verbundenen Metadaten. Diese Kosten entstehen auf der einen Seite bei jeder Organisation, die den DOI einsetzt, auf der anderen Seite aber auch bei der International DOI Foundation (bzw. der Registration Agency), die die Resolver-Datenbank aktuell halten muss. Die Höhe der jährlich (oder auch in einem anderen Zeitraum) anfallenden Gebühren für teilnehmende Organisationen ist bisher noch nicht abzusehen, könnte aber prohibitiv für kleinere Organisationen sein. Darüber hinaus führt die Nutzung des DOIs durch große Verlagsgesellschaften auch nicht zu mehr frei zugänglichen Informationen, da die meisten Objekte, die mit einem DOI versehen sind, nur mit den nötigen Zugangsrechten heruntergeladen werden können und so ebenso zu Frustration führen wie ein Error 404.

Im Gegensatz zum DOI, dessen Fortentwicklung von einem mächtigen Konsortium gefördert wird, verläuft die Umsetzung der URN-Infrastruktur sehr viel langsamer. Cuenca u. a. (1999) verweisen darauf, dass die technischen Anforderungen an Resolver-Dienste für URNs hinreichend beschrieben sind, bisher jedoch nur experimentelle Implementierungen verfügbar sind, wie z.B. das in Australien entwickelte BURNS (Basic URN Service resolution for the Internet) (vgl. Ianella u. a., 1996).

Das bisher einzige, in größerem Maßstab funktionierende System, ist das oben beschriebene Handle-System. Da es nicht auf den DOI als Identifier beschränkt ist,

²⁷<http://www.doi.org/idf-member-list.html>, Zugriffsdatum: 25.09.2000

kann es auch für andere Namespaces eingesetzt werden. Dies muss allerdings mit der Schaffung einer organisationalen Infrastruktur für die Vergabe und Verwaltung von Handles einhergehen.

4.4 Reference Linking und Citation Linking

Das herausragende Merkmal des World Wide Web ist die Verknüpfung von Informationseinheiten durch Hyperlinks. Für die wissenschaftliche Kommunikation ergibt sich durch die Verknüpfung von wissenschaftlichen Aufsätzen ein sehr viel schnellerer Zugang zu Informationen, als es mit gedruckten Zeitschriften möglich wäre. Hunter (1998) führt einige Beispiele für den sinnvollen Einsatz von Links in einer elektronischen Umgebung auf:

- von einer bibliographischen Suche zum Volltext eines Artikels
- vom Inhaltsverzeichnis zum Volltext
- von einer Literaturangabe zum Abstract oder Volltext des zitierten Artikels (Reference Linking)
- von einem Artikel zu späteren Zitierungen des Artikels
- von einer Zeitschrift eines Verlages zu einem anderen Verlag

Interessant für die weitere Betrachtung sind insbesondere die letzten drei Punkte: Welche Methoden gibt es, um (automatisiert) von einer Referenz zum Volltext des zitierten Artikels zu verweisen? Wie kann der Zugriff auf spätere Zitationen eines Artikels erreicht werden? und Wie kann dies verlagsübergreifend realisiert werden?

Im folgenden wird die Unterscheidung von de Solla Price (1986) für die Begriffe Referenz und Zitation verwendet:

[...] if Paper R contains a bibliographic footnote using and describing Paper C, then R contains a reference to C and C has a citation from R.

4.4.1 Grundannahmen für den Einsatz von Reference Linking

Nach Caplan und Arms (1999) ist *Reference Linking* ein allgemeiner Terminus für „links from one information object to another“. Mögliche Kontexte für Links sind dabei die Referenzen wissenschaftlicher Arbeiten, Referenzen von einem Bibliothekskatalog oder einer Bibliographie oder informelle Referenzen, die per Email oder verbal übermittelt werden.

Für die Domäne der Referenzen in Zeitschriftenartikeln lässt sich das Grundproblem des Reference Linking folgendermaßen in einer Frage formulieren: „Wenn die Informationen einer Zitierung in Zeitschriftenartikel gegeben sind, wie gelangt ein Nutzer von der Zitation zu einer Kopie dieses Artikels?“ Als grundlegendes Modell wurde im NISO Reference Linking Workshop angenommen: „Eine Zitation führt irgendwie zu einem Identifier, der irgendwie zu einer Kopie der zitierten Arbeit führt“ (Caplan, 1999). Abbildung 4.15 zeigt ein allgemeines Modell des *Reference Linking*. Wenn ein Nutzer auf eine Referenz stößt, kann er diese von einer *Reference Database* zu einem Identifier auflösen lassen. In der *Reference Database* sind Metadaten enthalten, die mit den Informationen in einer konventionellen Referenz korrespondieren. Ausgehend von den angefragten Daten werden alle Identifier zurückgegeben, deren Metadaten mit der Anfrage übereinstimmen. Der zweite Schritt ist die Auflösung des Identifiers zu einer URL mit Hilfe der *Location Database*. Die Auflösung von Identifiers zu URLs wurde in 4.3.4.4 und 4.3.5.6 beschrieben. Der letzte Schritt ist die Anforderung des tatsächlichen Inhalts mit der URL.

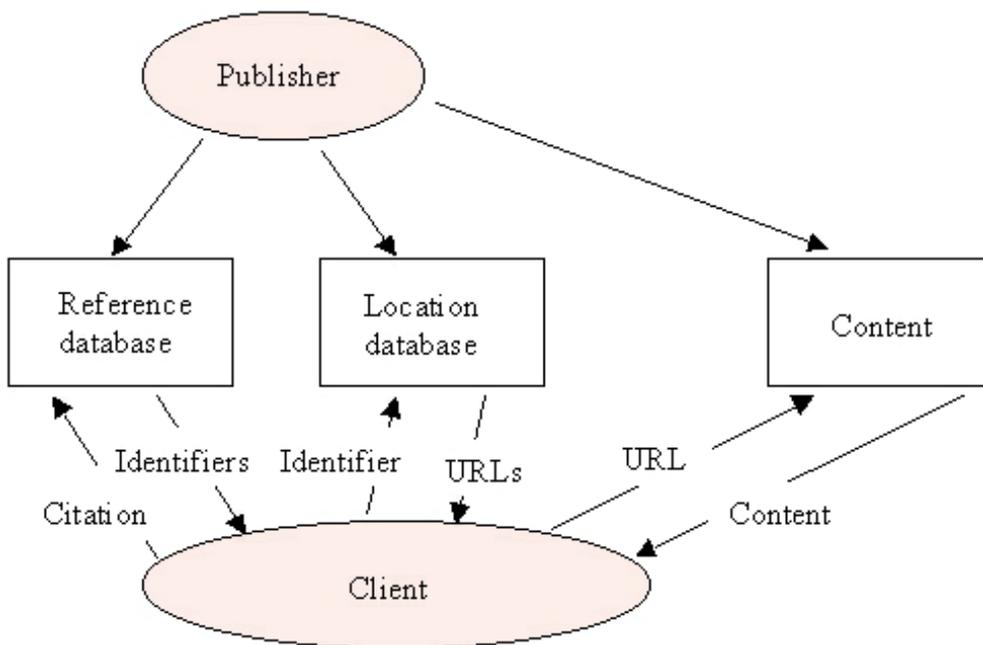


Abbildung 4.15: Allgemeines Modell des Reference Linking (Caplan und Arms, 1999)

Caplan und Arms (1999) weisen auf einige Schwierigkeiten und Sonderfälle hin, die bei diesem Prozess auftreten können. So ist die Form von Referenzen nicht standardisiert. Darüber hinaus kann der Prozess manchmal dadurch abgekürzt werden, dass ein Identifier in einer Referenz enthalten ist. Teilweise können mehrere Arbeiten in

der *Reference Database* mit den angefragten Daten übereinstimmen, und es muss eine Auswahl getroffen werden, entweder automatisch oder intellektuell. Ebenso kann ein Identifier zu mehreren URLs aufgelöst werden.

4.4.2 Auflösung einer Referenz zu einem Identifier

Nach Van de Sompel und Hochstenbach (1999a) lassen sich die Ansätze in der wissenschaftlichen Fachkommunikation zur Verlinkung von Aufsätzen in statische Verlinkung und dynamische Verlinkung unterscheiden.

4.4.2.1 Statische Verlinkung

Statische Verlinkung bedeutet, dass die Verknüpfungen zwischen Informationseinheiten bereits existieren und in einer Datenbank abgelegt sind. Hier gibt es drei unterschiedliche Möglichkeiten, um von einer Referenz zu einem Identifier zu gelangen (Caplan, 1999):

- In die Referenz ist ein Identifier eingebettet.
- Die bibliographischen Informationen in der Zitation können benutzt werden, um einen Identifier zu berechnen.
- Die bibliographischen Informationen in der Zitation können benutzt werden, um einen Identifier in einer Referenzdatenbank abzufragen.

Die erste Möglichkeit ist trivial, da hier nur ein Auflösungs-Prozess wie oben beschrieben angestoßen wird. Allerdings kann nicht davon ausgegangen werden, dass jede Referenz einen Identifier enthält.

Bei der Berechnung eines Identifiers aus bibliographischen Informationen wird aus den gegebenen Daten in der Referenz, wie Name der Publikation, Jahrgang und Seitenzahlen ein, eindeutiger Identifier abgeleitet. Ein bekanntes Beispiel für einen Identifier, der aus bibliographischen Informationen berechnet wird, ist der *Serial Item and Contribution Identifier (SICI)* (vgl. Nat, 1996). Der SICI besteht aus einem *Item Segment*, das die ISSN und Daten über Jahrgang, Ausgabe und Erscheinungsdatum beinhaltet, einem *Contribution Segment*, das eine Lokalisierungsangabe (z.B. Seitenzahlen) und einen Code für den Titel einer Arbeit enthält und einem *Control Segment*, das administrativen Zwecken dient. Der SICI ist mit hoher Wahrscheinlichkeit eindeutig, wenn alle grundlegenden bibliographischen Daten vorhanden sind.

Ein anderes Beispiel für die Anwendung eines berechneten Identifier ist das Schema der *American Physical Society*. Jedem Artikel wird ein sechsstelliger, numerischer Identifier zugewiesen, der aus den Komponenten *Ausgabe der Zeitschrift*, einer *Subjekt-Klassifikation* und der *Artikelnummer* besteht. Dieses Schema wird auch für elektronische Artikel angewendet, die vor der gedruckten Ausgabe der Zeitschrift im Internet publiziert werden. Mithilfe des *APS link manager* wird eine URL aus dem Identifier berechnet (Doyle, 2000).

Das Problem bei der Berechnung von Identifiers liegt in den verschiedenen Formen, in denen Verweise auftreten können. Die automatische Extraktion der entsprechenden Informationen aus einer Referenz ist daher ziemlich komplex. Ein gangbarer Weg ist die Anwendung sogenannter *Templates*, die mit den gebräuchlichen Verweisformaten verschiedener Verlage korrespondieren. Dennoch ist häufig eine zusätzliche intellektuelle Bearbeitung notwendig, da Literaturzitate häufig auch fehlerhaft sind. Das *Institute for Scientific Information (ISI)* und das *Scholarly Link Specification Framework* benutzen Templates zur Unterstützung bei Extraktion von Verweisdaten.

Die dritte Möglichkeit, um von einer Referenz zu einem Identifier zu gelangen, ist der Einsatz einer *Reference Database*. Hier werden die bibliographischen Informationen der Referenz nicht zur Konstruktion eines Identifier benutzt, sondern als Parameter für eine Datenbankabfrage. In diesem Zusammenhang stellt sich die Frage, welche Metadaten in der *Reference Database* notwendig sind. Basierend auf einem Vergleich verschiedener Systeme, wurde im NISO Reference Linking Workshop ein minimales Metadaten-Set vorgeschlagen (Caplan, 1999):

1. Title: Titel des Zeitschriftenartikels.
2. Creator(s): Autor des Artikels, wobei mindestens der erste Autor enthalten sein sollte.
3. Journal Title: Name der Zeitschrift, in welcher ein Artikel erschienen ist. Hier kann auch eine ISSN eingesetzt werden.
4. Date: Publikationsdatum des Artikels oder der Zeitschrift.
5. Enumeration: „Nummerierung“ der Zeitschriftenausgabe, z.B. Volume und Nummer der Ausgabe oder auch Teil einer Serie. Dies kann weggelassen werden, wenn eine Zeitschrift keine offizielle „Nummerierung“ hat.
6. Location: Seitenzahl der ersten Seite eines Artikels oder zugeordnete Artikelnummer, wenn keine Seitenzahlen vorhanden.

7. Type: Art des beschriebenen Materials, z.B. „Artikel in einer Fachzeitschrift“. Dieses Feld ist von Bedeutung, da es auch die Möglichkeit gibt, dass nicht zum Volltext eines Artikels verwiesen wird, sondern nur zu einem Abstract.

Diese Metadaten-Set ist kompatibel mit dem oben beschriebenen Metadaten-Set des Digital Object Identifier (siehe 4.3.5.5). Allerdings ist es schwierig, diese Elemente den entsprechenden Elementen in Dublin Core zuzuordnen, da in DC eine Möglichkeit fehlt, die Beziehung zwischen einer Fachzeitschrift und einem Artikel in einer Fachzeitschrift auszudrücken. In Apps und MacIntyre (2000) wird die Anwendung von Dublin Core auf bibliographische Daten beschrieben, wobei weitgehend der Empfehlung der *Bibliographic Citation Working Group* (vgl. Morgan, 1999) gefolgt wird. Das Problem der Zuordnung eines Artikels zu einer Ausgabe einer Zeitschrift wird dabei über den Qualifier `IsPartOf` für das Attribut `Relation` gelöst; der Wert dieses qualifizierten Attributes kann dann z.B. einen Identifier für eine bestimmte Ausgabe einer Zeitschrift enthalten (dies ist abhängig vom gegebenen SCHEME).

```
<DC.Identifier.Citation SCHEME="DCCITE">
  <JournalTitleFull>International Journal of Electrical
  Engineering Education</JournalTitleFull>
  <JournalTitleAbbreviated>IJEEE</JournalTitleAbbreviated>
  <JournalChronology>January 2000</JournalChronology>
  <JournalVolume>37</JournalVolume>
  <JournalIssueNumber>1</JournalIssueNumber>
  <JournalPages>26-37</JournalPages>
</DC.Identifier.Citation>

<DC.Identifier SCHEME="DOI">10.1060/IJEEE.2000.003
</DC.Identifier>

<DC.Relation.IsPartOf SCHEME="SICI">
0020-7209(20000101)37:1</DC.Relation.IsPartOf>
```

Abbildung 4.16: Kodierung von bibliographischen Informationen mit Dublin Core in XML (Apps und MacIntyre, 2000)

Eine allgemeinere Lösung für die Beschreibung der Beziehungen zwischen Zeitschriftenausgabe und Artikel könnte der Einsatz des Resource Description Framework sein.

4.4.2.2 Dynamische Verlinkung

Im Gegensatz zur statischen Verlinkung gibt es bei der dynamischen Verlinkung keine Datenbank mit vorausberechneten Verweisen. Links werden erst dann generiert, wenn

sie von einem Nutzer angefordert werden. Der Vorteil der dynamischen Verlinkung ist, dass nicht alle Daten, die für eine verlinkte Umgebung notwendig sind, unter der Kontrolle einer Organisation stehen müssen. Beispiele für die Generierung von Verweisen bei Anforderung sind das unten beschriebene SFX und der Distributed Link Service.

4.4.2.3 Adaption an die Umgebung

Van de Sompel und Hochstenbach (1999a) treffen eine Abgrenzung zwischen offenen und geschlossenen Systemen für die Verlinkung von Informationsobjekten. Geschlossene Systeme passen sich nicht an die Umgebung an, d.h. die Ziele der Links sind unabhängig von den Informationsressourcen der Institution. Da Institutionen wie Universitäten mit ihren Bibliotheken häufig verschiedene Informationsanbieter haben, führt dies zu einem Nebeneinander nicht-verlinkter Ressourcen.

Die möglichen Probleme geschlossener Link-Systeme werden in Van de Sompel und Hochstenbach (1999a) anhand einiger Beispiele verdeutlicht:

- Eine Institution möchte nicht, dass ein Link zu einer Pay-per-view-Version eines Artikels führt, weil es entweder ihrer Politik widerspricht oder weil der entsprechende Artikel lokal verfügbar ist.
- Eine Institution möchte dem Nutzer alternative Linkziele innerhalb eines für die Institution zugänglichen Bereiches anbieten. Dies betrifft beispielsweise auch Verweise zu verfügbaren Papierversionen von Aufsätzen, Verweise zum professionellen Hintergrund eines Autors (Indikator für Qualitätskontrolle), Verweise von einem Eintrag im OPAC zu einem Review oder Abstract des entsprechenden Buches.

Da diese Anwendungen in hohem Maße von der verfügbaren Infrastruktur abhängig sind, ist eine Integration dieser Ressourcen mit einem sogenannten geschlossenen System zur Verlinkung nicht zu erreichen. Mit SFX wird ein Beispiel für ein offenes, dynamisches System zur Verlinkung in 4.4.4.5 vorgestellt.

Caplan und Arms (1999) schlagen zwei mögliche Modelle für die selektive Auflösung (i.e. Linkziele abhängig vom Nutzerprofil) innerhalb geschlossener Systeme vor.

- Die in 4.4.1 beschriebene *Location Database* wird mit Informationen über den Nutzer angereichert. Die Institution hat so die Möglichkeit, die Selektion der Links, die einem Nutzer innerhalb der Institution präsentiert werden, zu steuern.

Beispielsweise wäre es dann möglich, dass ein Nutzer beim Folgen eines Links auf einen kostenpflichtigen Artikel, statt auf einen Pay-per-view-Hinweis, auf eine Seite der Bibliothek umgeleitet wird, die den Standort der Papierversion enthält.

Das Problem dieser zentralen Lösung liegt in der Skalierbarkeit. Da der Betreiber der *Location Database (Resolver Service)* Profile für viele Institutionen vorhalten muss, wird die benötigte Datenmenge sehr groß.

- Eine weitere Möglichkeit der selektiven Auflösung ist der Einsatz eines lokalen Resolvers. Wenn ein Nutzer einen Identifier auflösen möchte (z.B. einem Link folgt), dann wird zuerst von einem lokalen Resolver eine Abfrage gestartet, um den Identifier einer oder mehreren URLs zuzuordnen, die im Einflussbereich der Institution liegen. Wenn die lokale Auflösung fehlschlägt, dann wird der Identifier an eine globale *Location Database* weitergeleitet (vgl. 4.3.4.4).

Die Implementierung dieses Modells ist mit den heute verfügbaren Web-Browsern schwierig, da diese nicht für die Auflösung verschiedener Hierarchiestufen von Identifiers konfigurierbar sind. Als Lösung bietet sich hier der Einsatz von Proxy-Servern oder Browser-PlugIns wie im Handle-System an.

Bei beiden Modellen ist die Pflege eines Auflösungsprofils notwendig.

4.4.3 Citation Indexes und Citation Linking

Neben der einfachen Verlinkung von Referenzen mit den korrespondierenden Artikeln (oder auch Abstracts und anderen Datenquellen) ist eine fortgeschrittene Methode der Recherche das sogenannte *forward linking* oder *citation linking*. Für einen gegebenen Aufsatz werden alle Aufsätze angezeigt, die diesen zitieren. Abbildung 4.17 verdeutlicht das Prinzip: Ausgehend von dem grau eingefärbten Ausgangsdokument ist es zum einen möglich, auf die Artikel zuzugreifen, die das Ausgangsdokument zitiert (Dokumente links des Ausgangsdokuments), zum anderen kann man auf all diejenigen Dokumente zugreifen, die das Ausgangsdokument zitieren (Dokumente rechts des Ausgangsdokuments).

Die Idee der Citation Indexes wird in dem Artikel *Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas* von Garfield (1955) in der Zeitschrift *Science* beschrieben. Garfield nennt als Vorteile gegenüber herkömmlichen Klassifikationen und themenbasierten Indexen die „*cross-fertilization of subject fields*“, i.e. der Zugang zu Literatur aus gänzlich anderen Fachgebieten, und die

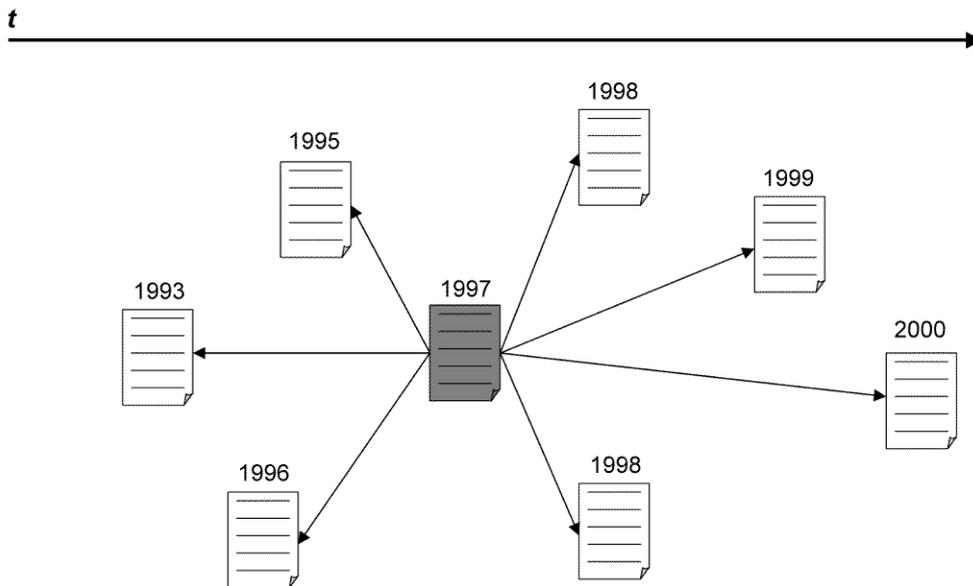


Abbildung 4.17: Citation Linking

Möglichkeit, den Ursprung einer Idee durch die Literatur zu verfolgen.

Darüber hinaus lassen sich mithilfe von Citation Indexes Zitationsanalysen durchführen, die als Ergebnis die Summe von Zitationen haben, die eine Fachzeitschrift, ein bestimmter Artikel oder gar ein Wissenschaftler über einen bestimmten Zeitraum erhalten hat. Eine bekannte Größe in diesem Zusammenhang ist der sogenannte *journal impact factor* des ISI, der ein Maß für die Bewertung von Fachzeitschriften darstellt²⁸ (vgl. Garfield 1994 für eine detaillierte Darstellung der Berechnung).

Die Citation Indexes wurden Anfang der 60er Jahre eingeführt und boten zuerst auf Papier, später auch in elektronischer Form die Möglichkeit der oben beschriebenen *cited reference search*. Atkins (1999) bezeichnet die Citation Indexes als „*prime example of a concept before its time*“, da erst durch die hypertextuellen Möglichkeiten des World Wide Web das Konzept für eine Vielzahl von Nutzern intuitiv zugänglich wurde. Die Implementierung des Citation Linking bei der ISI - *Web of Science* - wird unten genauer beschrieben.

4.4.4 Implementierungen und Werkzeuge

Im folgenden werden einige Implementierungen des Reference Linking und Citation Linking Konzepts genauer beschrieben.

²⁸Der tatsächliche Nutzen des Journal Impact Factor ist in der Literatur umstritten. Eine kritische Auseinandersetzung findet sich beispielsweise bei Seglen (1997) und Stegmann (1999).

4.4.4.1 Web of Science

Seit 1998 bietet das ISI mit dem Web of Science ihre Citation Indexes (kostenpflichtig) auch über das World Wide Web an. Der Prozess der Erfassung wird in Atkins (1999) beschrieben. Jede Zeitschrift, die bei der ISI eingeht, wird *cover-to-cover* erfasst, d.h. die erfassten Objekte umfassen Artikel, Leserbriefe, Reviews, Korrekturen etc.. Aus den bibliographischen Daten dieser Objekte wird dann ein eindeutiger Bezeichner generiert. Darüber hinaus werden die Referenzen erfasst und die Bezeichner für die Referenzen generiert. Anschließend wird eine Datenbankabfrage gestartet, die nach den Bezeichnern für die Referenzen in der Datenbank sucht. Wird ein Bezeichner gefunden, dann besteht ein Link zwischen zwei Objekten. Falls kein Bezeichner gefunden wird, dann werden die bibliographischen Informationen der Referenz strukturiert erfasst. In einem späteren Schritt wird genauer analysiert, ob das Objekt eventuell doch erfasst ist und nur die Angaben der Referenz fehlerhaft sind.

Dieses standardisierte Verfahren der Erfassung und Vereinigung der Referenzen hat nach Meinung von Atkins drei Vorteile:

1. Die Datenerfassung kann sehr schnell geschehen, da die Bezeichner sehr kurz sind.
2. Die Referenzen können für den Nutzer in einem standardisierten Format dargestellt werden.
3. Interne und externe Links können in ISI Produkte integriert werden.

Interne Links im Web of Science Von einem Eintrag für einen Artikel im Web of Science gibt es drei unterschiedliche Arten von internen Links:

Reference Links Von den Referenzen in einem Artikeleintrag kann der Nutzer zu den entsprechenden Einträgen in der Datenbank navigieren.

„Times Cited“Links Im Gegensatz zu den *Reference Links* bieten *„Times Cited“Links* die Möglichkeit, zu denjenigen Artikeln zu navigieren, die das Ausgangsobjekt zitieren (siehe auch Abbildung 4.17).

Related Records Links Diese Links ermöglichen die Navigation zu Artikeln, die mit dem Ausgangsartikel einen oder mehrere Referenzen gemeinsam haben (*bibliographic coupling*).

Externe Links im Web of Science Wie oben beschrieben, bietet das Web of Science vielfältige Möglichkeiten der Navigation zwischen Datenbankeinträgen von Artikeln (oder anderen Objekten). Der nächste logische Schritt ist die Verknüpfung dieser Einträge zu anderen elektronisch vorliegenden Produkten. Derzeit existieren Links vom Web of Science zu drei verschiedenen Arten von externen Informationsquellen:

Full-Text Links Von den Referenzen eines Eintrags im Web of Science existieren Links zu den entsprechenden Volltexten der Verlage.

Die Generierung der Links zu den Volltexten der Verlage wird folgendermaßen durchgeführt: Der betreffende Verlag sendet für jeden Artikel bibliographische Daten, einen eindeutigen Identifier und die URL an das ISI. Aus den bibliographischen Daten wird dann nach demselben Verfahren wie bei der Generierung interner Bezeichner ein Schlüssel erzeugt. Überall dort, wo die Schlüssel für die externen Quellen mit internen Bezeichnern übereinstimmen, wird ein Link zwischen ISI-Eintrag und externer Quelle eingefügt.

Patent Links Patentdaten aus dem *Derwent Innovations Index* werden auch mit dem Web of Science verlinkt. Zum einen werden die Referenzen in den Patentschriften mit den entsprechenden Einträgen für die Artikel verlinkt, zum anderen werden die Referenzen in Artikeleinträgen mit den entsprechenden Patentschriften verknüpft.

Sequence Data Links In Kooperation mit der National Library of Medicine's National Center for Biotechnology Information (NCBI)²⁹ wird das Web of Science mit GenBank³⁰ verlinkt.

Die Verlinkung der Volltexte mit den entsprechenden Einträgen im Web of Science kann bei Zeitschriften Probleme machen, die bestimmte Artikel elektronisch vor Erscheinen der gedruckten Version zugänglich machen, weil Informationen zur Artikelidentifikation, wie z.B. die Seitenzahlen, fehlen. Dies macht die algorithmische Generierung von Links unmöglich³¹.

²⁹<http://www.ncbi.nlm.nih.gov>, Zugriffsdatum: 06.10.2000

³⁰GenBank ist eine Datenbank mit Gensequenzen (<http://www.ncbi.nlm.nih.gov/Genbank/>), Zugriffsdatum: 14.11.2000

³¹Springer oder Wiley identifizieren Artikel, die elektronisch, aber noch nicht gedruckt erschienen sind, mit dem DOI.

4.4.4.2 Reference Linking mit DOIs (CrossRef)

Der Einsatz des Digital Object Identifier für die Verlinkung von Referenzen mit den korrespondierenden Artikeln ist eine offensichtliche Anwendung für einen Identifier. Anfang 1999 wurde die Entwicklung eines Prototypen von der IDF begonnen, und im Oktober 1999 wurde auf der Frankfurter Buchmesse ein funktionierender Prototyp (DOI-X) vorgestellt (Atkins u. a., 2000). Seit Februar 2000 wird die Entwicklung von CrossRef getragen, einem Konsortium bestehend aus 58 wissenschaftlichen Verlagen³²; seit dem 6. Juni 2000 ist das System bei den teilnehmenden Verlagen in Betrieb³³.

Da die CrossRef-Initiative auf die technische Dokumentation des DOI-X Prototypen verweist, wird bei der Beschreibung der zugrundeliegenden Infrastruktur nur noch von DOI-X die Rede sein.

Im Rahmen des DOI-X Projektes wurden folgende Komponenten für das Reference Linking mit DOIs entwickelt (Lyons und Ratner, 1999, S. 2):

- Eine XML DTD und ein Regelset für den Upload von DOIs und den dazugehörigen Metadaten auf einen zentralen Datenbankserver.
- Abfragen für diese Datenbank im Batch-Verfahren: Die resultierenden DOIs können dann von einem Informationsanbieter in die Referenzen eingefügt werden.
- Interaktive Abfragen für die Datenbank: Diese können z.B. von Autoren wissenschaftlicher Aufsätze benutzt werden, um DOIs direkt in ihre Referenzen einzufügen.
- Eine Offline-Lieferung der Datenbank an Archivdienste, damit diese die entsprechenden DOIs in ihre Referenzen einfügen können.

Die Funktionsweise entspricht dem oben beschriebenen, allgemeinen Modell von Caplan und Arms (siehe Abb. 4.4.1); das Metadatenmodell wurde bereits in 4.3.5.5 beschrieben. Der Upload der Metadaten und die Registrierung der DOIs mit den korrespondierenden URLs ist in Abb. 4.18 schematisch dargestellt.

Abbildung 4.19 zeigt den Vorgang der Abfrage der DOI-X Datenbank mit Informationen zu der Ressource. In der Datenbank können mehrere DOIs für eine bestimmte Ressource existieren. Gegenwärtig sind diese beschränkt auf die Typen *abstract only*,

³²<http://www.crossref.org/members.htm>, Zugriffsdatum: 02.10.2000

³³<http://www.crossref.org/faqs.htm>, Zugriffsdatum: 02.10.2000

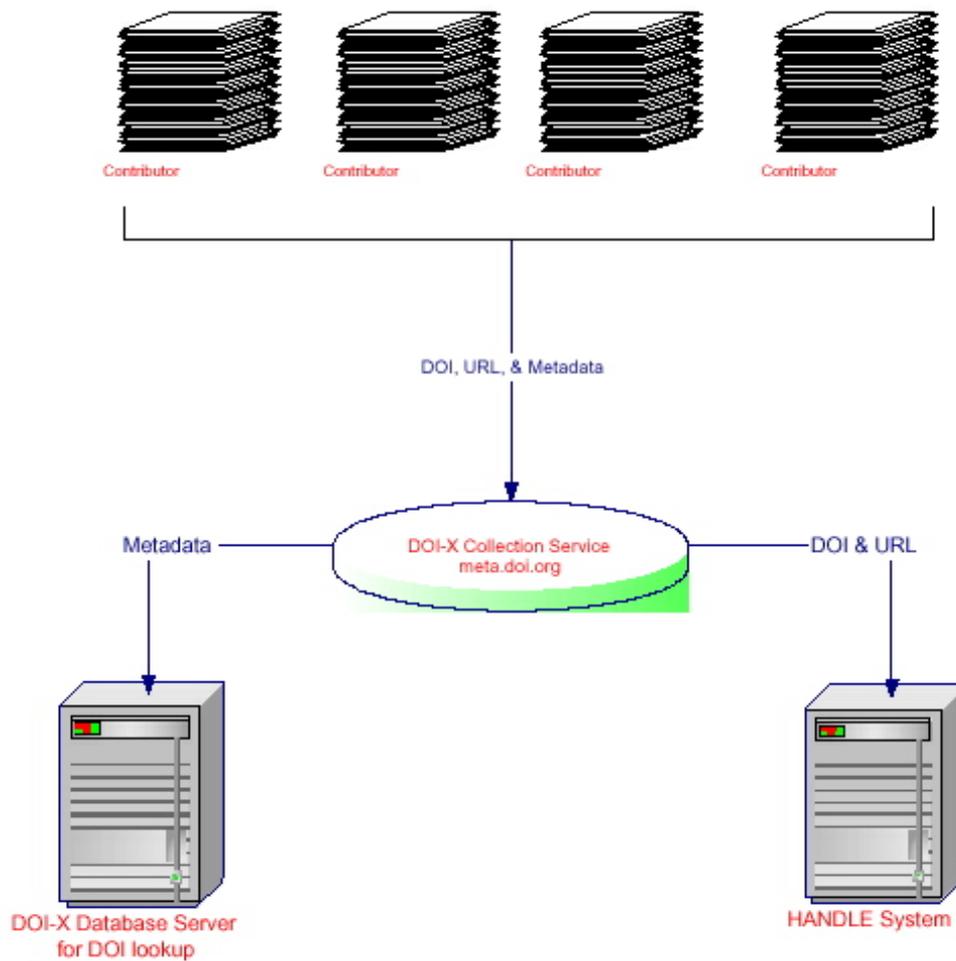


Abbildung 4.18: Sammlung von Metadaten und DOI/URL-Registrierung

full text und *biblio*. Der abfragende Dienst kann die zurückgelieferten DOIs dann als Hypertext-Link in seine Referenzen einfügen.

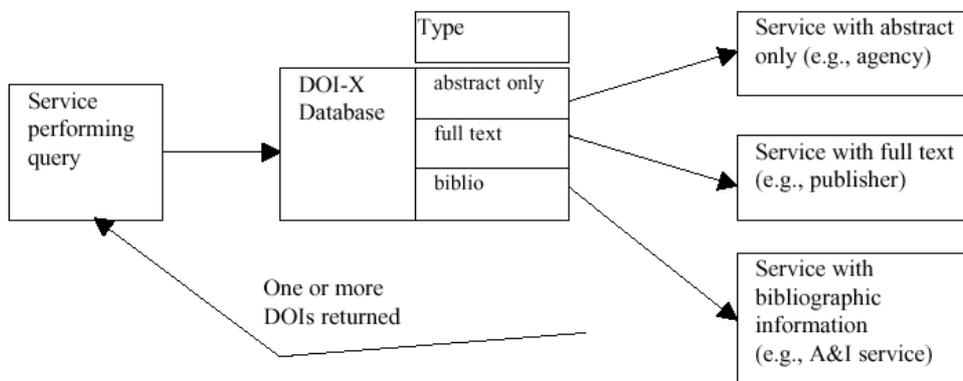


Abbildung 4.19: Abfrage der DOI-X Datenbank

In Lyons und Ratner (1999) werden drei mögliche Nutzergruppen für die Metadaten-Datenbank angenommen:

Contributors Dies sind Verlage und andere Informationsanbieter, die Links zu ihren Produkten ermöglichen möchten. Sie registrieren DOIs mit den entsprechenden Metadaten bei CrossRef.

Service-level users Hierzu gehören Abstracting- und Index-Dienste, Verlage etc., die die Referenzen ihrer Produkte mit Hypertext-Links zu Volltexten, Abstracts oder bibliographischen Einträgen anreichern möchten.

Research-level user Dies sind Autoren, die ihre Referenzen mit Hypertext-Links zu Volltexten, Abstracts oder bibliographischen Einträgen anreichern möchten.

Die Nutzung von CrossRef ist kostenpflichtig für die beiden ersten Gruppen. Verlage bezahlen einen jährlichen Mitgliedsbeitrag, der abhängig von der Zahl der registrierten Artikel ist. Darüberhinaus wird pro registriertem Artikel eine Gebühr von 0.60US\$ erhoben und pro gefundener DOI in der Metadaten-Datenbank ein Betrag von 0.05US\$³⁴.

4.4.4.3 SLinkS

Das *Scholarly Link Specification Framework (SLinkS)* bietet ein Vokabular und eine Syntax für den Austausch von Linking-Informationen zwischen verschiedenen Organisationen. SLinkS besteht dabei aus den Komponenten *URL templating language* und einem *metadata vocabulary*. Die *URL templating language* gibt vor, wie eine URL aus bibliographischen Daten konstruiert werden kann. Dies geschieht mit bibliographischen Platzhaltern in der URL-Zeichenkette. Ein Beispiel für die Konstruktion von URLs zeigt Tabelle 4.6.

<i>Volume</i>	<i>Startpage</i>	<i>URL</i>
3	25	http://www.publisher.com/003/25/
10	485	http://www.publisher.com/010/485/

Tabelle 4.6: Beispiel für die URL einer Fachzeitschrift unter <http://www.publisher.com>, basierend auf dem Volume und der Startseite des Artikels (Hellman, 1999b, S. 70)

Für die Verlinkung von Referenzen ist die Existenz von Metadaten notwendig, die eine Ressource hinsichtlich Namen, Identifiers und Abkürzungen, ihrer Art und ihrer

³⁴Details zu den Gebühren von CrossRef finden sich unter <http://www.crossref.org/docs/fees.htm>, Zugriffsdatum: 02.10.2000

Stabilität beschreiben. SLinkS verfügt daher über ein eigenes Metadaten-Vokabular, das in RDF (siehe 4.2.4) formuliert ist.

Hellman (1999a) beschreibt drei verschiedene Szenarios, in denen SLinkS zum Einsatz kommen kann. Beim ersten Szenario (Abbildung 4.20) wird von einer zentralen Clearing-Stelle ausgegangen, die die SLinkS-Spezifikationen sammelt und pflegt (im Beispiel von Verlag C und D). Wenn beispielsweise Verlag A die Referenzen seiner Produkte verlinken möchte, sendet er die Referenzdaten zum SLinkS-Linker. Der Linker benutzt die Datenbank mit den Spezifikationen zur Generierung von URLs zu den Quellen und sendet diese zurück an Verlag A. Dieser kann die Links dann seinen Nutzern zur Verfügung stellen.

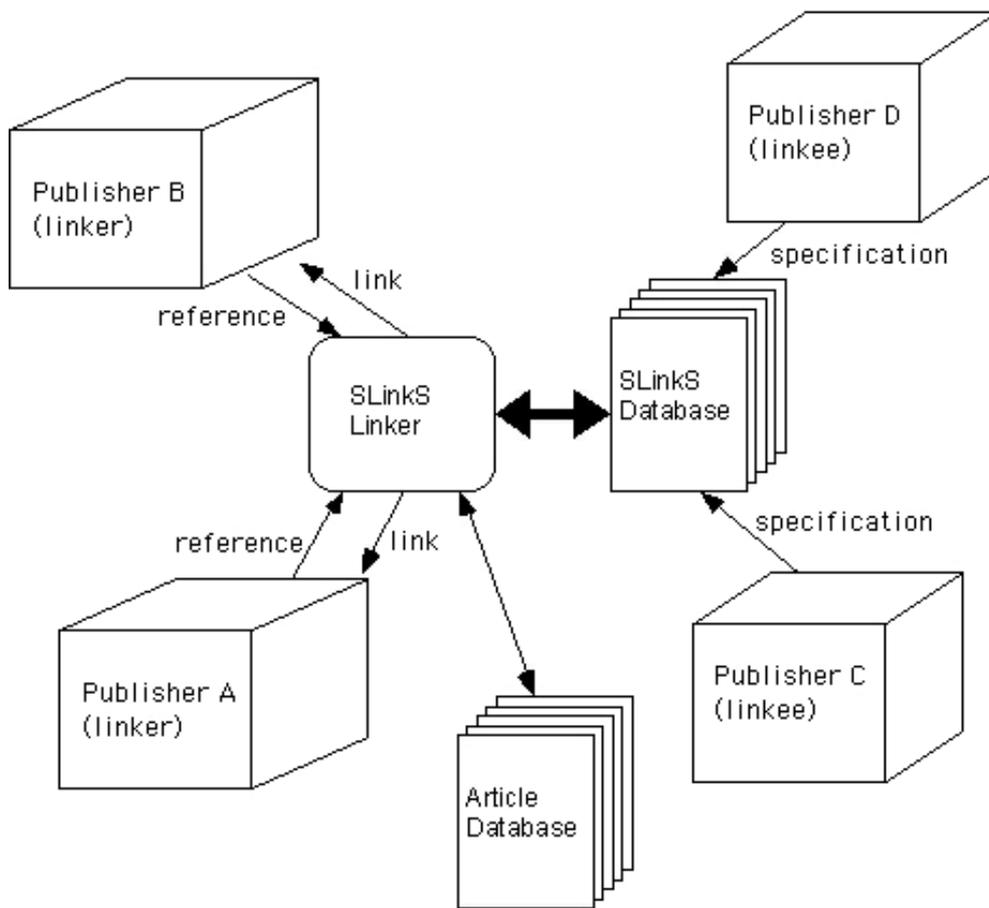


Abbildung 4.20: Zentralisierte Linking-Architektur

Bei dem zweiten Szenario (Abbildung 4.21) hat ein Verlag A eine Datenbank mit SLinkS-Spezifikationen in seinem Einflussbereich. In der Datenbank sind auch Spezifikationen eines anderen Verlages abgelegt (Verlag C). Die Inhalte, die einem Nutzer

zur Verfügung gestellt werden, sind dynamisch aus der Content Database generiert. Der SLinkS-Linker konstruiert die Links bei der Generierung der angeforderten Seite. Die dynamische Generierung ermöglicht eine beschränkte zeitliche Gültigkeit von Links. Dies ist insbesondere dann von Bedeutung, wenn Inhalte von anderen (hier Verlag C) betroffen sind.

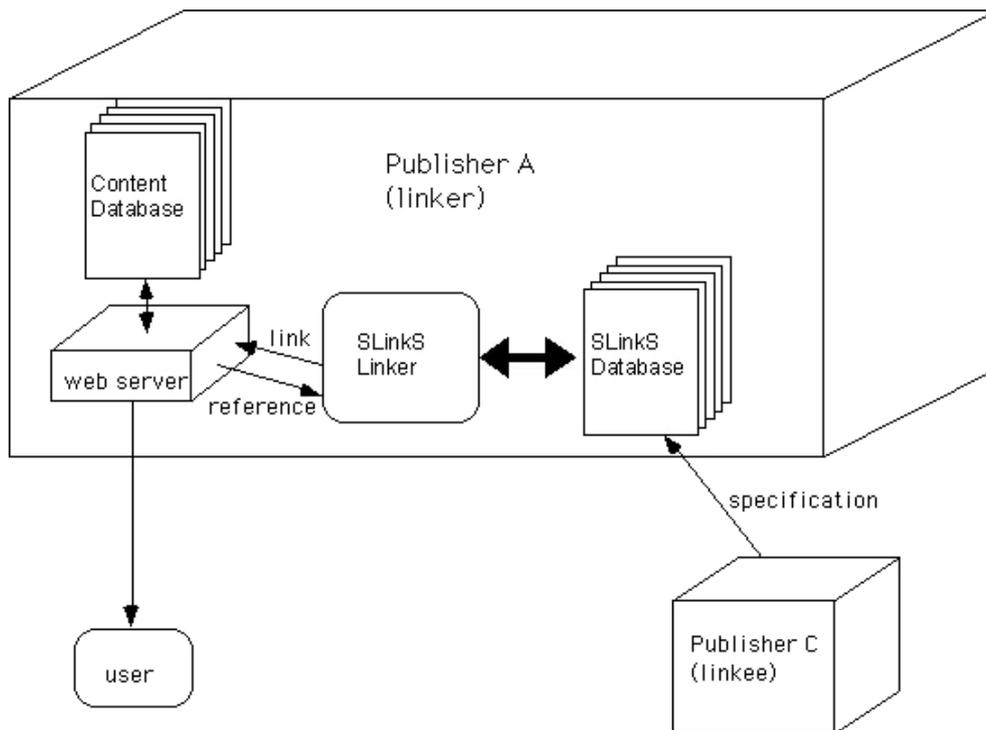


Abbildung 4.21: Private Linking-Architektur

Das dritte Szenario geht von der Sicht eines Bibliotheksbenutzers aus. Die Bibliothek hat von dem Verlag B den Volltext einer Zeitschrift abonniert. Ein Nutzer folgt einem Link, den der Verlag A zu einem Artikel in einer Zeitschrift des Verlages B zur Verfügung gestellt hat. Der SLinkS Proxy Server erkennt, dass der Volltext der Zeitschrift für Bibliotheksbenutzer zugänglich ist und leitet den Nutzer entweder zu einer lokalen Kopie des Artikels oder zu einer Kopie auf dem Server von Verlag B.

Gegenwärtig existieren schon URL-Templates für viele Verlage, die ihre Zeitschriften über das Internet anbieten³⁵. Das Geschäftsmodell von Openly Informatics sieht vor, dass der Zugang zur Spezifikationen-Datenbank und zu dem JAVA-basierenden Metadaten-Werkzeugen kostenfrei ist. Die Software zur Umsetzung der Metadaten-Spezifikationen in die URLs ist kostenpflichtig.

³⁵<http://www.openly.com/SLinkS/faq.html>, Zugriffsdatum: 04.10.2000

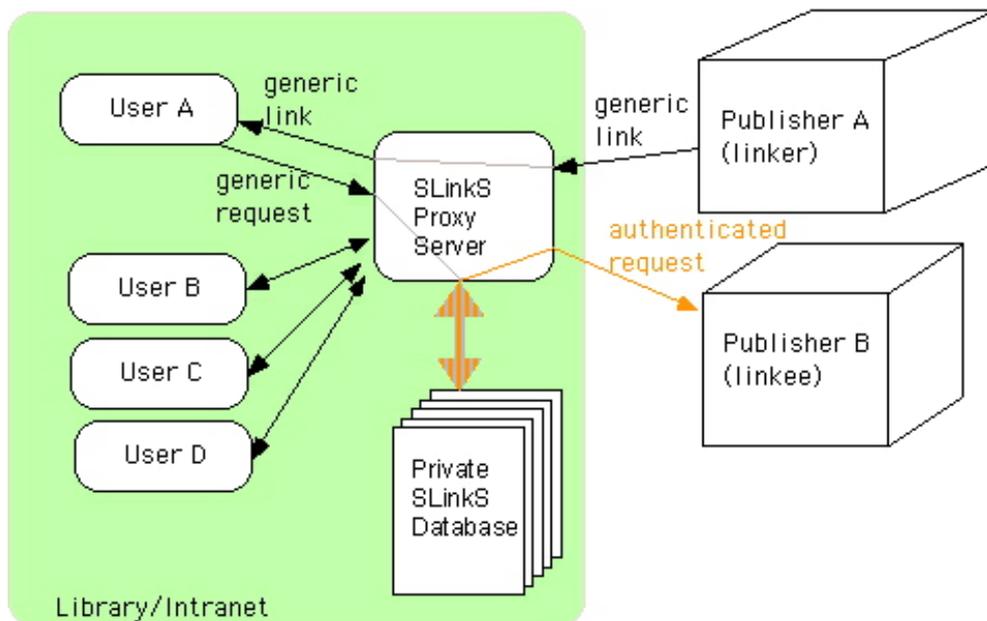


Abbildung 4.22: Intranet Linking-Architektur

4.4.4.4 Research Index

Einen interessanten Ansatz für den automatisierten Aufbau eines unabhängigen Zitierindex beschreiben Lawrence u. a. (1999a). Das *Autonomous Citation Indexing System* kann für Literatur in elektronischer Form einen Zitierindex generieren. Ein solches System muss in der Lage sein, wissenschaftliche Aufsätze im Internet zu lokalisieren, die Referenzen (mit ihrem Kontext) zu extrahieren und Referenzen auf gleiche Artikel in unterschiedlichen Formaten zu erkennen.

Lokalisierung von Dokumenten Die Lokalisierung von Aufsätzen kann auf mehreren Wegen geschehen:

- Artikel können mit Hilfe von Suchmaschinen und geeigneten Heuristiken gefunden werden (z.B. Seiten, die Begriffe wie „publications“, „paper“ oder „post-script“ enthalten)
- Mailinglisten können für Verweise auf Aufsätze gescannt werden
- Die Web-Seiten von Verlagen können untersucht werden. Da der Zugriff auf die Volltexte von Fachzeitschriften meist kostenpflichtig ist, sind hier Vereinbarungen mit den Verlagen notwendig
- Autoren benachrichtigen das System selbst

Der Prototyp eines solchen Systems - *CiteSeer* - lädt PostScript- oder PDF-Dateien herunter, die zur Verarbeitung in normalen ASCII-Text konvertiert werden. Die Existenz eines Literaturverzeichnis wird überprüft, um festzustellen, ob es sich tatsächlich um einen wissenschaftlichen Aufsatz handelt.

Verarbeitung der Dokumente Der erste Schritt in der Verarbeitung der Dokumente ist die Lokalisierung des Literaturverzeichnisses und die Extraktion von Autor, Titel, Jahr, Seitenzahlen und Schlüssel (z.B. „Müller, 1982“) aus den Referenzen. Diese Extraktion basiert auf Heuristiken hinsichtlich des Aussehens von Referenzen. Beispielsweise gehen Informationen über den Autor normalerweise den Titelinformationen voraus. Das größte Problem bei der Verarbeitung ist die Identifizierung von Referenzen, die auf den gleichen Aufsatz verweisen, da das Aussehen der Referenzen oft sehr unterschiedlich ist. *CiteSeer* benutzt hier einen Algorithmus, der auf der Normalisierung der Referenzen basiert mit einer Sortierung hinsichtlich Wortlänge und gefundenen Wörtern und Phrasen in den Subfeldern.

Eine Demoversion - der *NEC ResearchIndex*³⁶ (vormals *CiteSeer*) - kann kostenfrei genutzt werden und bietet Zugriff auf elektronische Aufsätze im Internet über eine Volltextsuche, eine Zitationsuche und eine Suche nach *related* und *overlapping* Dokumente. Darüber hinaus bietet das System die Erstellung von Zitationsgraphen und eine Alert-Funktion, wenn ein bestimmter Artikel eine weitere Zitation erhält. Der Quelltext des Systems ist für den nicht-kommerziellen Gebrauch frei verfügbar (Lawrence u. a., 1999b).

4.4.4.5 SFX

Einen dynamischer Ansatz zur Verlinkung heterogener Informationsressourcen beschreiben Van de Sompel und Hochstenbach (1999b). SFX bietet für Bibliotheken die Möglichkeit der Kombination von internen mit externen Ressourcen, z.B. kann der Verweis auf ein Buch in einem Literaturverzeichnis eines Aufsatzes zu dem entsprechenden Eintrag im OPAC der Bibliothek oder zu der entsprechenden Seite bei einem Internet-Buchhändler verlinkt werden. Im Gegensatz zu den festverdrahteten Links zwischen Identifiern und URLs bei anderen Ansätzen, verfügt SFX nur über eine Datenbank mit *Conceptual Links* zwischen verschiedenen Informationsangeboten.

Wenn einem Nutzer ein Informationsangebot präsentiert wird, sind potenzielle Link-Quellen mit einem SFX-Button versehen. Hinter dem SFX-Button verbirgt sich

³⁶<http://www.researchindex.com>, Zugriffsdatum: 05.10.2000

ein Identifier, der die ID des Servers der Linkquelle, die ID der Datenbank, in der die Link-Quelle abgelegt ist, die ID des Eintrags der Link-Quelle in der Datenbank und den SFX-Prozess, der bei Klick auf den Button ausgeführt werden soll, enthält. Der SFX-Button enthält noch keine Information über das Ziel, auf das verlinkt werden soll, sondern nur Informationen über die Link-Quelle. Der nächste Schritt ist die Umwandlung dieser Informationen in das SFX-Format. In der *SFX Local Service Component* werden die Parameter der Link-Quelle mit den *Conceptual Links* abgeglichen. Alle gültigen Links zwischen Informationsangeboten werden dem Nutzer angezeigt, d.h. der Nutzer kann aus unterschiedlichen Informationsangeboten wählen. Wenn der Nutzer sich für ein bestimmtes Angebot entschieden hat und auf den entsprechenden Verweis klickt, wird ein *Target Parser* gestartet, der die in dem Link kodierten Metadaten in die entsprechende URL für die Ressource umwandelt.

SFX wurde ursprünglich im Automation Department der Zentralbibliothek der Universität von Ghent in Belgien entwickelt (Van de Sompel und Hochstenbach, 1999a). In einem großangelegten Experiment am Los Alamos National Laboratory wurde die Anwendbarkeit auf eine komplexe Informationsinfrastruktur gezeigt (Van de Sompel und Hochstenbach, 1999c). Seit Februar 2000 wird es als kommerzielles Produkt von Ex Libris³⁷ vertrieben und weiterentwickelt.

4.4.4.6 Distributed Link Service

Der *Distributed Link Service* (Carr u. a., 1995), der an der Universität von Southampton entwickelt wurde, verfolgte ursprünglich einen breiteren Ansatz als „nur“ die Verlinkung von Referenzen. Ziel war die Schaffung eines vom spezifischen Dokument-unabhängigen Link-Dienstes, der Verknüpfungen zwischen Dokumenten dynamisch mit Hilfe einer Abfrage von Link-Datenbanken generiert. Der DLS ermöglicht dem Nutzer, in einem Text verschiedenen Wörter zu markieren und diese als Abfrage an eine oder mehrere Link-Datenbanken zu senden. Der DLS enthält mehrere Datenbanken:

document-specific linkbase In dieser Datenbank sind alle Dokument-spezifischen Verweise abgelegt.

resource-specific linkbase Enthält Verweise für ein Gruppe von Dokumenten.

server-specific linkbase Enthält Verweise für alle Dokumente, die auf einem Server abgelegt sind.

³⁷<http://www.exlibris-usa.com>, Zugriffssdatum: 05.10.2000

user-specific linkbase Enthält die Verweise, die von dem Nutzer angelegt wurden, der die Anfrage gestartet hat.

context-specific linkbase Enthält Verweise, die während der Ausführung einer bestimmten Aufgabe relevant sind.

Neben den verschiedenen Link-Datenbanken und der Komponente zum Suchen in den Datenbanken besteht das System noch aus einer Client-Komponente, die die Formulierung von Anfragen an die Link-Datenbanken ermöglicht.

Im *Open Journal Project*, das von 1995 bis 1998 innerhalb des *UK Electronic Library (eLib)* Programmes unterstützt wurde, sollten die Möglichkeiten der Anwendung von Links in der wissenschaftlichen Kommunikation untersucht werden (Hitchcock, 1998). Zu diesem Zweck setzte man die Technologie des *Distributed Link Service* für die Verlinkung von Volltexten ein. Wenn ein Dokument angefordert wurde, erkannte ein sogenannter *citation agent* die enthaltenen Referenzen, suchte nach Treffern in einer Datenbank und verlinkte diese mit den im Artikel enthaltenen Referenzen (Hitchcock u. a., 1998).

4.4.4.7 Andere Ansätze

Neben den beschriebenen verlagsübergreifenden Bestrebungen zur Verlinkung von Volltexten haben einige Verlage und Organisationen für ihre Publikationen diese Technik implementiert, wie z.B. die American Physical Society³⁸, die NASA mit dem Astrophysics Data System³⁹ oder das PubMed/PubRef-System der National Library of Medicine⁴⁰.

4.4.5 Diskussion

Die beschriebenen Ansätze unterscheiden sich hinsichtlich ihrer Ausrichtung sehr stark:

- Das Web of Science des Institute of Scientific Information stellt ein leistungsfähiges Werkzeug für die Recherche nach Aufsätzen dar. Im Vordergrund steht hier insbesondere das Citation Linking. Der Zugriff auf Volltexte über Abkommen mit Fachverlagen und die Verlinkung mit externen Datenbanken runden das System ab. Das System ist als solches geschlossen, d.h. es können nur

³⁸<http://www.aps.org>, Zugriffsdatum: 12.11.2000

³⁹<http://adswwww.harvard.edu/>, Zugriffsdatum: 12.11.2000

⁴⁰<http://www.nih.gov/welcome/director/pubmedcentral/ebiomedarch.htm>, Zugriffsdatum: 11.11.2000

Artikel von denjenigen Fachzeitschriften recherchiert werden, die vom ISI erfasst werden.

- Die CrossRef-Initiative legt ihren Schwerpunkt auf das Reference Linking und den Zugriff auf Volltexte. Problematisch hierbei ist die Fokussierung auf die großen Verlage, die ihre elektronischen Artikel mit einem DOI versehen. Darüber hinaus sind die Kosten, die für die Registrierung eines Artikels und die Auflösung einer Referenz entstehen, nicht unerheblich. Weiterhin ist das System nicht an lokale Gegebenheiten anpassbar.
- SLinkS bietet weniger eine fertige Lösung für das Reference Linking sondern vielmehr Werkzeuge, mit denen sich ein an die individuelle Infrastruktur einer Institution angepasstes Reference Linking-System aufbauen lässt.
- Der bei NEC entwickelte ResearchIndex zeigt, wie vollkommen automatisiert Links zwischen elektronisch verfügbaren Artikeln aufgebaut werden können. Der ResearchIndex eröffnet dabei ähnliche Recherchemöglichkeiten wie das Web of Science (Citation Links, Zitationsanalyse etc.); bedingt durch die rein maschinelle Verarbeitung der Dokumente wird allerdings nicht die Qualität der Ansätze erreicht, die Metadaten für einzelne Aufsätze erfassen.
- Das mittlerweile kommerzielle SFX ist kein Reference Linking oder Citation Linking System im Sinne der vorher vorgestellten Ansätze, sondern bietet die Integration und Verlinkung aller in einer Institution vorgehaltenen Informationsressourcen unter einer Oberfläche.
- Einen vom Ziel her ähnlichen Ansatz wie SFX verfolgt das Distributed Link System (DLS). Auch hier können unterschiedlichste Ressourcen miteinander verlinkt werden, wobei es sowohl vom System (d.h. von den einzelnen Datenbanken) vorgegebene, als auch vom Nutzer zu generierende Links gibt.

4.5 Management des Zugangs zu digitalen Objekten

Die hier vorgestellten Ansätze des Zugangsmanagements gehen von einer Umgebung aus, in der eine Institution ihren Mitgliedern Zugang zu lizenzierten digitalen Objekten ermöglichen möchte. Ein typisches Beispiel für eine solche Umgebung ist eine Universität, deren Universitätsbibliothek ihren Nutzern Zugang zu ihren abonnierten elektronischen Zeitschriften gewähren möchte.

Bisher bieten nur sehr wenige Verlage⁴¹ ein *Pay-per-view* an, d.h. man bezahlt nur für die Artikel, die man sich auch tatsächlich herunterlädt. Die breite Akzeptanz für *Pay-per-view* scheitert bisher noch aus zwei Gründen: Zum einen sind bisher sichere Micropayment-Verfahren im Internet noch nicht etabliert (siehe Sietmann 2000b und Wasmeier 2000 für einen Überblick über Online-Zahlungsmittel), zum anderen möchten die wissenschaftlichen Verlage die Abonnements der Bibliotheken als Kalkulationsgrundlage nicht aufgeben; *Pay-per-view* kann hier höchstens ein Zusatzangebot für Abonnenten sein (vgl. Sietmann, 2000a).

4.5.1 Allgemeines Modell der Autorisierung

Abbildung 4.23 zeigt ein Modell für das Zugangsmanagement. Eine Organisation (z.B. eine Bibliothek) erzeugt Richtlinien für den Zugang zu digitalen Materialien, basierend auf bestehenden Gesetzen und Vereinbarungen (z.B. mit Verlagen oder Datenbankanbietern). Diese Richtlinien verbinden Benutzer und digitale Materialien. Benutzer müssen authentifiziert werden, und ihre Rollen für den Zugriff auf die Ressourcen müssen festgelegt werden. Rollen sind z.B. die Zugehörigkeit zu einer Gruppe (Student oder Mitarbeiter der Universität des Saarlandes) oder der Ort, von dem aus ein Benutzer auf die Informationen zugreifen möchte (vgl. 4.5.2.1). Das digitale Material muss identifiziert werden, i.e. jedes Objekt benötigt einen eindeutigen Identifier. Für bestimmte Domänen ist die Authentifizierung von Ressourcen von Bedeutung, d.h. die Versicherung, dass ein bestimmtes Objekt nicht absichtlich oder versehentlich geändert wurde. Dies ist beispielsweise bei Verträgen u. ä. von großer Bedeutung (vgl. Arms, 2000, S. 123-131). Die Authentizität elektronischer Dokumente kann durch digitale Signaturen oder auch durch das sogenannte *watermarking*, i.e. die unsichtbare Einbettung von Urheberrechtshinweisen, sichergestellt werden (Lynch, 1996, S. 140/1).

Wenn ein Benutzer Zugang zu elektronischen Ressourcen haben möchte, wird ein

⁴¹Ein Beispiel ist die Association of Computing Machinery (ACM). Diese bietet sowohl Mitgliedern ihrer Organisation als auch Nichtmitgliedern die Möglichkeit des *Pay-per-view* für einzelne Artikel an.

Prozess angestoßen, in dessen Verlauf die Identität eines Nutzers überprüft wird und eine Autorisierungsprozedur ihm erlaubt oder verbietet, bestimmte Operationen mit den digitalen Materialien auszuführen.

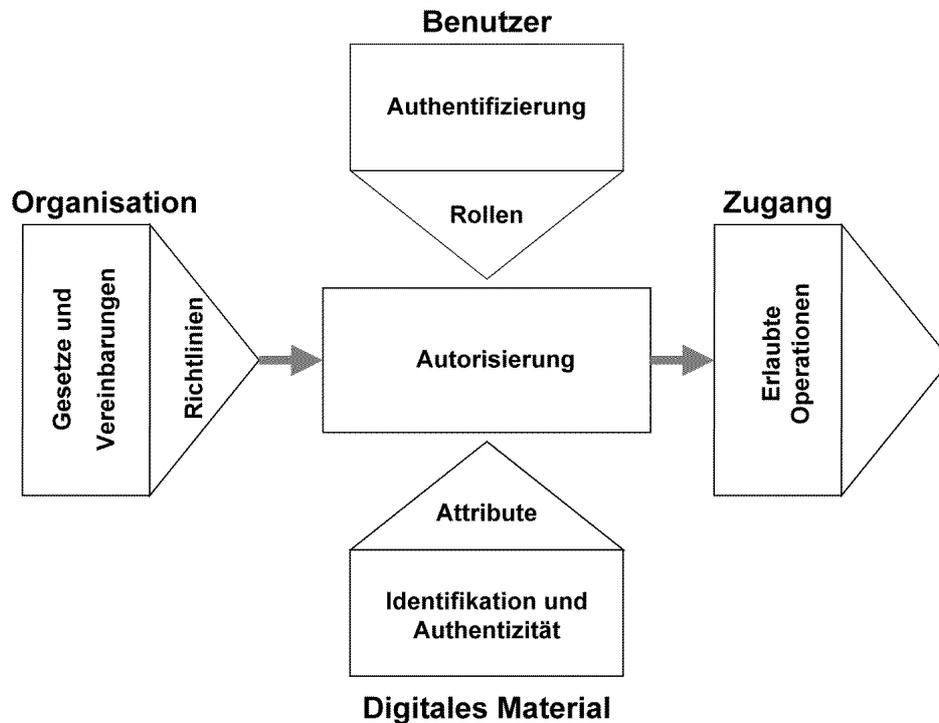


Abbildung 4.23: Modell des Zugangs-Managements (vereinfacht nach Arms, 1999a, S.34)

4.5.2 Technologien für das Zugangsmanagement

Die folgende Aufstellung von Technologien für das Zugangsmanagement ist von Lynch (1998) und Krieb (1999) abgeleitet.

4.5.2.1 IP-Filter

Die am wenigsten aufwendige Form der Autorisierung ist die Überprüfung der IP-Adresse, von der eine Anfrage kommt. Ein Verlag führt auf seinen Dokumenten-Servern eine Datenbank mit erlaubten IP-Adressen und/oder -Bereichen. Ist eine IP innerhalb eines verzeichneten Bereiches, dann ist die Anfrage autorisiert und wird ausgeführt. Diese Form der Autorisierung wird von sehr vielen Verlagen aufgrund der niedrigen Wartungskosten gewählt.

Das Hauptproblem dieser Methode ist, dass viele Mitglieder einer Institution nicht unbedingt durch die Institution selbst Zugang zum Internet haben. Statt des Zugangs

über das Rechenzentrum ihrer Universität oder Institution benutzen sie einen kommerziellen Internet Service Provider und können - weil ihr Provider einen anderen Adress-Bereich als ihre Organisation benutzt - somit keinen Zugang zu diesen autorisierungspflichtigen Angeboten erhalten. Darüber hinaus ist durch die rechnergebundene Art des Zugangsmanagements nur eine grobkörnige Vergabe von Zugangsprivilegien möglich. Dies trifft insbesondere dort zu, wo sich mehrere Personen einen Rechnerarbeitsplatz teilen (z.B. CIP-Pools an Universitäten).

4.5.2.2 Benutzername/Passwort

Eine andere Möglichkeit, den Zugang zu kostenpflichtigen Angeboten zu kontrollieren, ist der Einsatz von Benutzernamen und Passwörtern. Um eine bestimmte elektronische Zeitschrift benutzen zu können, muss ein Nutzer einen Benutzernamen und ein Passwort eingeben. Wenn diese Kombination vom Informationsanbieter als gültig überprüft wurde, erhält der Nutzer Zugang zu dem entsprechenden Angebot. Neben der Vergabe von persönlichen Benutzer-Accounts bieten einige Anbieter institutionsweite Kennungen und Passwörter an.

Die Methode der Überprüfung von Benutzer-Accounts hat den Vorteil, dass ein Benutzer unabhängig vom Ort auf die gewünschten Ressourcen zugreifen kann. Problematisch ist hingegen der Verwaltungsaufwand: Für Institutionen mit einer großen Nutzerfluktuation, wie z.B. Universitäten, müssen die Listen mit den Benutzer-Accounts ständig gepflegt werden, da neue Nutzer hinzukommen und alte gelöscht werden müssen. Die Vergabe von institutionsweiten Kennungen bedeutet zwar sehr viel weniger Pflegeaufwand, auf der anderen Seite besteht hier eher die Gefahr des Missbrauchs von Zugangsrechten. Ein weiteres Problem ergibt sich dadurch, dass Institutionen die Produkte von einer Vielzahl von Anbietern lizensieren. Dies führt zu einer großen Anzahl von Passwörtern, die von Nutzern gemerkt werden müssen.

4.5.2.3 Credential-Based

Credential-based bedeutet, dass der Zugang zu Ressourcen über die Referenzen des Nutzers (z.B. Zertifikate) gewährt wird. In der Praxis erhält der Benutzer ein elektronisches Zertifikat, das seine Identität und seine Zugehörigkeit zu einer bestimmten Gruppe bestätigt. Der Vorteil dieser Zertifikate ist, dass ein bestimmter Nutzer unabhängig vom Ort über eine *certificate authority* eindeutig authentifiziert werden kann. Der gegenwärtig bekannteste Standard ist X.509. Eine detaillierte Beschreibung von X.509 und ein Vergleich mit anderen kryptographischen Zertifizierungsverfahren kann

in Gerck (2000) nachgelesen werden.

Der weitflächige Einsatz von X.509 wird behindert durch die notwendige Infrastruktur und die Komplexität der Verteilung der Zertifikate und Schlüssel zum Nutzer. Darüber hinaus erschweren international unterschiedliche Bestimmungen hinsichtlich der Benutzung und des Exportes von kryptographischen Verfahren die Verbreitung.

4.5.2.4 Proxy-Server

Eine vielversprechende Technologie sind Proxy-Server zur Regelung des Zugangs zu bestimmten Ressourcen. Anders als IP-Filter ist hier eine Unabhängigkeit des Nutzers von einem Ort gegeben, da die Authentifizierung über den Proxy-Server abläuft. Der Aufwand für einen Benutzer ist relativ gering: Sie/Er muss einmal in den Einstellungen ihres/seines Browsers die URL des entsprechenden Proxy-Servers und von der Institution zugeteilte Benutzernamen und Passwort eintragen. Auf der Institutionsseite ist die Verwaltung von Benutzeraccounts notwendig. Da hier die Informationsanbieter nicht betroffen sind, hat eine Institution eine große Flexibilität hinsichtlich der Zugangsgewährung zu den abonnierten Produkten innerhalb ihres Mitgliederkreises. Ein Nachteil dieser Lösung ist die Angreifbarkeit des Proxy-Server als zentrales Authentifizierungssystem. Bei einem Ausfall sind für alle Nutzer die angebotenen Dienste gesperrt. Krieb (1999) weist daraufhin, dass bestimmte Verlage den Zugang zu ihren Produkten über Proxy-Server nicht erlauben.

4.5.3 Diskussion der vorgestellten Ansätze

Von den vier vorgestellten Ansätzen haben insbesondere die letzten beiden das Potenzial, die Anforderungen, die von Nutzern und auch Informationsanbietern an Systeme zum Zugangsmanagement gestellt werden, zu erfüllen. Das Filtern von IP-Adressen ist zwar eine kostengünstige Möglichkeit des Zugangsmanagements, erlegt den Nutzern allerdings zu viele Beschränkungen auf. Der Zugang über Benutzername/Passwort ist insbesondere bei großen Institutionen mit mehreren Lizenzen für Informationsprodukte problematisch, da zum einen der Verwaltungsaufwand für die Vielzahl von Benutzern immens ist und zum anderen für die Benutzer nicht sehr komfortabel ist, da sie sich an eine große Anzahl Passwörter erinnern müssen.

Die Autorisierung über einen Proxy-Server oder über Zertifikate scheinen vielversprechende Lösungen für das Problem des Zugangsmanagements zu sein, da hier zum einen Benutzer direkt authentifiziert werden können und nicht nur ein Rechner und zum zweiten der Nutzer sich nur eine Kennung für den Zugriff auf verschiedene

Dienste merken muss.

Ein wichtiger Aspekt des Zugangsmanagement ist die Sicherstellung der Privatsphäre der Nutzer (vgl. Arms, 2000, S. 119). In der elektronischen Welt ist es für Informationsanbieter sehr einfach, die Nutzung seiner Angebote zu verfolgen (welches Material wird wie oft abgerufen?). Wenn sich Nutzer über Zertifikate oder direkt über eine persönliche Kennung und Passwort authentifizieren, können die Nutzungsdaten mit Individuen in Zusammenhang gebracht werden, was der Politik der Bibliotheken zuwiderläuft (Lynch, 1998).

Eine komfortable und sichere Möglichkeit der Benutzer-Authentifizierung stellen biometrische Verfahren dar, die Nutzer über physische Attribute wie Fingerabdruck, Retina, Stimme oder auch das Gesicht identifizieren. Insbesondere bei der Zutrittskontrolle zu Räumen oder Gebäuden sind diese Verfahren bereits akzeptiert, im Bereich des Access Managements wirken die Investitionen für die Zusatzhardware bisher hemmend (Diedrich, 1999).

4.6 Weitere Aspekte

4.6.1 Usability

Ein wichtiger Punkt für die Akzeptanz einer digitalen Infrastruktur ist die *Usability* der Zugangssysteme zu digitalen Kollektionen. In diesem Zusammenhang bedeutet dies beispielsweise, dass das Oberflächendesign konsistent ist, die Bedienung der Kontrollen klar ersichtlich ist und der Nutzer eine Rückmeldung seiner Aktionen erhält. Darüber hinaus sind kurze Antwortzeiten des Systems ein wesentlicher Faktor für die Nutzerzufriedenheit (Arms, 2000, S. 144ff.).

Ausgehend von einer Usability-Studie existierender digitaler Bibliotheken konstruierten Kengeri u. a. (1999) eine Taxonomie der Charakteristika, die für ein digitales Bibliothekssystem essentiell sind. Tabelle 4.6.1 zeigt die Charakteristika und ihren Ausprägungsgrad bei den untersuchten digitalen Bibliotheken.

4.6.2 Retro-Digitalisierung

Die Mehrzahl der Fachzeitschriften der großen wissenschaftlichen Verlage erscheint heute sowohl in gedruckter als auch in elektronischer Form (siehe 3.1.2.1). Viele wichtige ältere Publikationen sind allerdings nur in gedruckter Form verfügbar, d.h. es ergibt sich ein Medienbruch innerhalb der oben beschriebenen Infrastruktur.

Die konsequenteste Lösung für dieses Problem ist die sogenannte Retro-Digitalisierung, i.e. die elektronische Abtastung mithilfe eines Scanners. Im einfach-

Features	ACM	IEEE-CS	NCSTRL	NDLTD	DeLiver
Clear Overview	•	•			••
Search criteria for simple search	••	••	••	••	••
Search criteria for advanced search	••		•		••
Fast searching and retrieval	•	••	••	••	•
Example searches	•	•		••	••
Ability to download a fraction of the article				••	
Save queries for future Refinement					••

- - Library has robust coverage of feature
- - Library supports the functionality minimally
- Library does not support functionality

Tabelle 4.7: Taxonomy of features for designing digital libraries (Kengeri u. a., 1999)

sten Fall werden die gescannten Seiten als Rasterbilder abgespeichert; eine andere Möglichkeit ist die Umwandlung der Rasterbilder in ein Textformat. Die dazu nötige OCR-Software (Optical Character Recognition) hat inzwischen zwar eine hohe Qualität erreicht, dennoch muss in Abhängigkeit von der Qualität der Vorlage häufig manuell nachgebessert werden.

Diese manuelle Nachbesserung ist auch der Grund für die immensen Kosten, die durch Retro-Digitalisierung entstehen: Die *SIGs* (*Special Interest Groups*) der ACM müssen bei der Erfassung älterer Konferenzbände umgerechnet ca. 8 DM für Schwarzweiss- und ca. 16 DM für Farbseiten aufwenden (Endres und Fellner, 2000, S. 285). Odlyzko (1997) berichtet von ähnlichen Kosten innerhalb des *JSTOR*⁴²-Projekts, wobei insbesondere die Klassifikation und Vollständigkeitsüberprüfung der gescannten Zeitschriftenausgaben die hohen Kosten verursachen.

⁴²JSTOR wurde 1995 ins Leben gerufen und versorgt Bibliotheken auf elektronischem Wege mit zurückliegenden Jahrgängen wissenschaftlicher Fachzeitschriften. Zwei Ziele stehen dabei im Vordergrund: Zum einen das akademische Ziel des Aufbaus eines Archivs und des Zugang zu diesem, und zum zweiten das wirtschaftliche Ziel, eine Kostenersparnis für die Bibliotheken zu erreichen, da nicht mehr jede Bibliothek das gleiche Material vorhalten muss (Arms, 2000, S. 55/6).

Kapitel 5

Andere Ansätze für die wissenschaftliche Kommunikation über das Internet

Die in Kapitel 4 beschriebene Infrastruktur stellt die formelle Komponente der wissenschaftlichen Kommunikation in den Vordergrund. Neben der Umsetzung des vorhandenen Publikationssystems in die elektronische Welt existieren noch weitere Ansätze, die zum einen erst durch die elektronische Kommunikation ermöglicht werden und zum anderen sehr stark auf eine Kombination von formeller und informeller Kommunikation abzielen.

5.1 Modulare Struktur für elektronische wissenschaftliche Artikel

Ein interessanter Ansatz für elektronische Aufsätze wird von Harmsz u. a. (2000) vorgeschlagen. Basierend auf der Gliederungsstruktur von Aufsätzen in der Physik, wird ein nichtlinearer, modularer Aufbau für elektronische Aufsätze vorgeschlagen. Folgende Module werden dabei unterschieden:

METAINFORMATION enthält bibliographische Angaben, Abstract, Inhaltsverzeichnis, Danksagung und Referenzen

POSITIONING beschreibt den Kontext der Arbeit und die zentrale Problemstellung

METHODS beschreibt, welche experimentellen, numerischen und theoretischen Methoden verwendet wurden

RESULTS enthält die Primärdaten und die vom Autor aufbereiteten Daten

INTERPRETATION enthält die quantitativen und qualitativen Ergebnisse

OUTCOME beschreibt die Antwort auf die zentrale Problemstellung und einen Ausblick auf weitere Forschungsarbeiten basierend auf den Ergebnissen

Die einzelnen Module desselben Aufsatzes und verschiedener Aufsätze werden dabei durch Links mit vorgegebenem Typ verknüpft. Zwei Klassen von Links werden unterschieden, *organisational relations* und *scientific discourse relations*. *Organisational relations* geben formale Beziehungen wie Hierarchien, verwandte Themengebiete, unterschiedliche Repräsentationen gleicher Sachverhalte etc. an, mit *scientific discourse relations* können Autoren die Beziehungen zwischen Modulen anderer Arbeiten knüpfen, wobei hier die Klassen *Relations based on communicative functions* und *Content relations* unterschieden werden. Die erste Klasse beinhaltet Beziehungen wie *wird erklärt in*, *wird definiert von* oder *wird bestritten von*, die zweite Klasse beinhaltet Beziehungen wie *wird genauer beschrieben in*, *wird verglichen mit* etc..

Bisher existiert nur ein Demonstrator, der zwei themenverwandte Aufsätze enthält¹. Werkzeuge zu einer großflächigen Implementierung fehlen bisher. Basierend auf den Erfahrungen mit dem Demonstrator bedeutet die modulare Struktur insbesondere, dass Module von Autoren wiederverwendet werden können und somit die Effizienz des Autors erhöhen. Darüber hinaus soll die Typisierung der Module und Links das Information Retrieval verbessern und Lesern eine selektive Auswahl der für sie relevanten Module ermöglichen. Die modulare Struktur kann desweiteren zu verständlicheren Publikationen führen.

5.2 Die Modelle von Hurd

Ausgehend von Garvey/Griffith Modell der wissenschaftlichen Kommunikation 2.1 hat Hurd (1996) einige Modelle für die zukünftige elektronische Kommunikation entwickelt. Von Interesse sind dabei insbesondere das „Unvetted Model“ und das „Collaboratory Model“, da diese einen besonderen Fokus auf die informelle Kommunikation legen und nur in einer vernetzten Umgebung existieren können.

5.2.1 The Unvetted Model

Dieses Modell (Hurd, 1996, S. 27/8) verzichtet auf ein Peer Review. Aufsätze werden auf einem öffentlich zugänglichen Server abgelegt und können von Interessierten

¹<http://www.wins.uva.nl/projects/commphys/papers/demo/demo.html>,
Zugriffsdatum: 30.10.2000

kommentiert werden. Die Anmerkungen und Kritikpunkte werden im weiteren Verlauf von dem Autor des Artikels aufgenommen und fließen in eine überarbeitete Fassung des Aufsatzes ein. Alle Revisionen des Aufsatzes und Kommentare zu diesem Aufsatz werden archiviert und bleiben zugänglich.

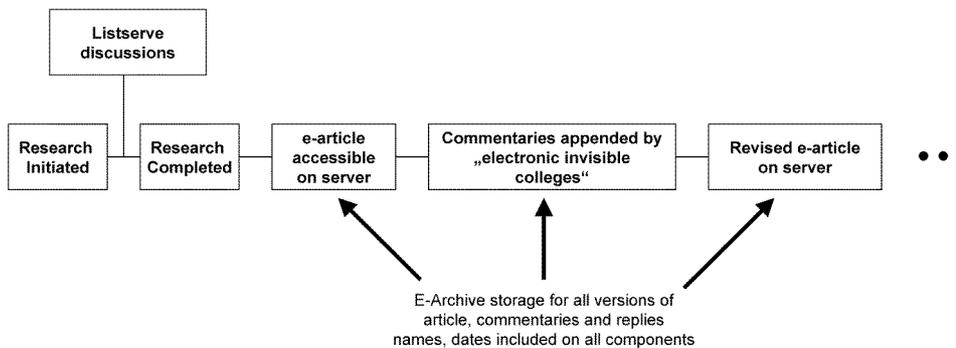


Abbildung 5.1: The Unvetted Model (Hurd, 1996, S. 28)

5.2.2 The Collaboratory Model

Im *Collaboratory Model* ist eine Publikation nicht mehr zentraler Bestandteil der wissenschaftlichen Kommunikation, sondern dient nur noch zum Festhalten von Resultaten. Der Austausch zwischen Wissenschaftlern findet in diesem Modell über elektronische Diskussionen und Datenbanken statt, in die gefundene Ergebnisse abgelegt werden und von anderen Wissenschaftlern direkt kommentiert und bewertet werden können. Nach Abschluss einer Forschungsarbeit werden die Daten zusammengefasst und elektronisch in Form eines Forschungsberichts verfügbar gemacht.

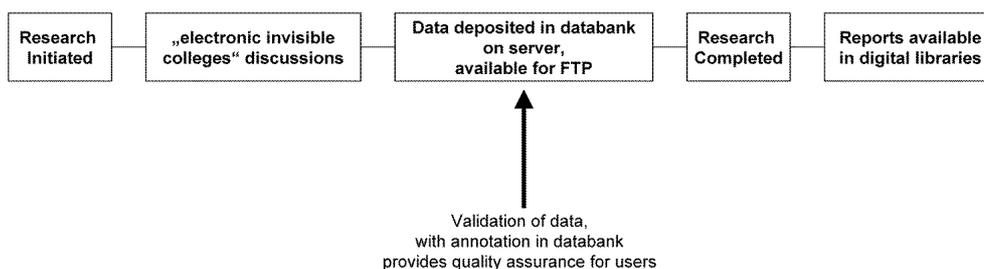


Abbildung 5.2: The Collaboratory Model (Hurd, 1996, S. 30)

Diese Art der Kommunikation eignet sich insbesondere für Großforschungsprojekte, wie beispielsweise in der Astronomie oder der Genetik. Ein bekanntes Beispiel für die Arbeit in diesem *Collaboratory Model* ist das Human Genome Project², das die menschlichen Gen-Sequenzen kartiert.

²<http://www.ornl.gov/hgmis/>, Zugriffsdatum: 06.11.2000

5.3 Scholarly Skywriting

Der Terminus *Scholarly Skywriting* wurde von Harnad (1991) geprägt und bezeichnet die elektronische Diskussion von wissenschaftlichen Fachartikeln. Das Scholarly Skywriting basiert auf Erfahrungen mit der psychologischen Zeitschrift *Behavioural & Brain Sciences (BBS)*. Ein wesentliches Merkmal von BBS war eine sogenannte Peer Commentary: Nachdem ein eingereichtes Manuskript einem Peer Review unterzogen wurde, konnte (mit Zustimmung des Autors) das Manuskript an interessierte Wissenschaftler gesendet werden, die den Aufsatz mit 1000 Worten kommentieren konnten. Der akzeptierte Aufsatz wurde dann zusammen mit diesen Kommentaren abgedruckt. BBS war mit diesem Modell zwar sehr erfolgreich (gemessen an seinem „Impact Factor“), dennoch war es der geringen Geschwindigkeit des Mediums Papier unterworfen.

Die elektronische Zeitschrift PSYCOLOQUY³ setzt das Konzept von BBS in die elektronische Kommunikation um. Auf PSYCOLOQUY kann kostenfrei zugegriffen werden, entweder über das World Wide Web, per ftp, per Email-Abonnement oder über die USENET-Newsgroup `sci.psychology.journals.psycology`. Im elektronischen Medium ist es nun möglich, innerhalb von Minuten die Beiträge von anderen Wissenschaftlern zu kommentieren. Harnad sieht das „electronic skywriting“ als die vierte Revolution in der Evolution der menschlichen Kommunikation und Kognition, neben Sprache, Schrift und Druck.

³<http://www.cogsci.soton.ac.uk/psycology/>, Zugriffsdatum: 30.10.2000

Kapitel 6

Zusammenfassung und Fazit

6.1 Zusammenfassung

Das traditionelle print-basierte Publikationssystem leidet insbesondere unter dem starken Wachstum der zu publizierenden wissenschaftlichen Literatur und den daraus resultierenden Verzögerungen bei der Veröffentlichung von Aufsätzen. Ein weiteres Problem ergibt sich durch die Preisgestaltung der wissenschaftlichen Verlage und der Finanzknappheit der Bibliotheken, was zu Problemen in der wissenschaftlichen Literaturversorgung führt.

Die formelle wissenschaftliche Kommunikation über das Internet bietet Vorteile hinsichtlich der Publikationsgeschwindigkeit, des Komforts für den Nutzer und Mehrwertfunktionen, wie Einbindung von Primärdaten und Multimediaelementen. Der weitflächigeren Nutzung dieses Medium stehen aber noch einige Probleme gegenüber, wie die Qualität von Suchmaschinentreffern, die eindeutige und dauerhafte Identifizierung und Lokalisierung von Objekten im Internet, der Problemkomplex der Interoperabilität und das Management des Zugangs zu digitalen Kollektionen.

Ausgehend von den Vorteilen und den Problemfeldern des elektronischen Publizierens in der Wissenschaftskommunikation wurde ein Szenario entworfen, das den formellen wissenschaftlichen Austausch über das Internet aus Nutzersicht illustriert und die Anforderungen an eine Infrastruktur definiert.

Das Retrieval von elektronischen Dokumenten im Internet kann durch den Einsatz von Metadaten signifikant verbessert werden. Im Bereich der deskriptiven Metadaten existiert mit Dublin Core ein stabiler Standard für die Beschreibung von Web-Ressourcen. In Verbindung mit Metadata-Registries und Crosswalks lässt sich die Interoperabilität von Metadaten-Sets unterschiedlicher Anwendungen realisieren. Das vom W3-Consortium vorgeschlagene Resource Description Framework (RDF) bie-

tet zum einen eine leistungsfähige, strukturierte Syntax für Metadaten, zum anderen können durch den Namespace-Mechanismus verschiedene Metadaten-Sets nebeneinander eingesetzt werden. RDF ist ein wesentliches Element des „Semantic Web“, der universellen, globalen Datenbank.

Für die dauerhafte Identifizierung von Web-Ressourcen gibt es mit der URN ein universelles Schema, das allerdings bisher noch nicht in großem Stil implementiert wurde. Der von der Verlagsindustrie aufgebrachte Digital Object Identifier (DOI) greift etwas weiter, da hier neben der persistenten Identifizierung auch kommerzielle Anwendungen im Blickpunkt stehen. Auf Basis der Handle-Technologie existiert bereits ein funktionierendes System zur Auflösung von DOIs in URLs, und das System wird bei allen wichtigen wissenschaftlichen Verlagen eingesetzt. Die Kosten für die Vergabe von DOIs und der Einfluss der Großverlage könnten sich als ein potenzielles Hemmnis für die weite Verbreitung des DOI erweisen.

Bei der Verlinkung von wissenschaftlichen Artikeln ist zu unterscheiden zwischen statischer und dynamischer Verlinkung: Statische Verfahren setzen eine Datenbank mit allen zu verlinkenden Informationseinheiten ein, dynamische Verfahren generieren aus den bibliographischen Daten einer Referenz eine URL. Neben diesen grundsätzlichen Verfahren zur Erzeugung der Links ist die Einbettung in die vorhandene Informations-Infrastruktur ein wichtiger Aspekt. Ein geschlossenes System, wie das auf dem DOI basierende CrossRef, ermöglicht keine lokale Filterung der enthaltenen Identifiers und ermöglicht daher nur den Zugriff auf globale Ressourcen. Adaptive Systeme wie SFX bieten der einsetzenden Institution vielfältige Konfigurationsmöglichkeiten hinsichtlich der Steuerung des Zugangs zu Ressourcen. Eine Sonderstellung nimmt das Web of Science der ISI ein: Das System bietet keine direkte Verlinkung von Volltexten, sondern nur die Verlinkung von bibliographischen Einträgen, wobei sowohl Reference Linking als auch Forward Linking unterstützt wird. Über Vereinbarungen mit wissenschaftlichen Verlagen kann auf die Volltexte von Zeitschriftenartikeln zugegriffen werden. Ein System zur automatischen Verlinkung von im Internet verfügbaren Aufsätzen ist der frei verfügbare ResearchIndex von NEC. Dokumente werden indexiert und die enthaltenen Literaturangaben werden verarbeitet und mit in der Datenbank enthaltenen Dokumenten verknüpft. Die Qualität der Verlinkung bleibt allerdings hinter der manuellen Verlinkung des Web of Science zurück.

Die Beschränkung des Zugangs zu digitalen Kollektionen ist insbesondere für das kommerzielle, wissenschaftliche Verlagswesen von großer Bedeutung. Für große Institutionen wie Universitäten ist die Überprüfung der IP-Adresse des anfragenden Rech-

ners das gegenwärtig am weitesten verbreitete Verfahren. Wenn diese Adresse bei einem Anbieter registriert ist, wird der Zugang zu allen abonnierten Materialien gewährt. Diese Methode ist allerdings sehr inflexibel, da nur Anfragen aus dem Netz der Institution autorisiert werden können, d.h. es werden keine Benutzer autorisiert, sondern nur Computer. Mittelfristig bietet die Autorisierung über institutionseigene Proxy-Server und über Nutzer-Zertifikate die besten Perspektiven für ein effektives Zugangsmanagement. Die Autorisierung über Proxy-Server ermöglicht einer Institution eine sehr feinkörnige Vergabe von Zugangsrechten und ist unabhängig von Rechneradressen. Eine Autorisierung über Zertifikate bietet die gleichen Möglichkeiten, die Handhabung ist aber sowohl für Nutzer als auch Anbieter komplexer. Ein wichtiger Aspekt des Zugangsmanagement ist der Schutz der Privatsphäre der einzelnen Benutzer; dies muss beim Design einer Infrastruktur mitberücksichtigt werden. Langfristig könnten auch biometrische Verfahren zur Benutzer-Authentifizierung eingesetzt werden.

Andere Ansätze für die wissenschaftliche Kommunikation betonen insbesondere die informelle Komponente und sind ohne elektronische Kommunikationsmöglichkeiten nicht implementierbar.

6.2 Fazit

In der Arbeit wurde aufgezeigt, dass die Technologien für eine effektive wissenschaftliche Kommunikation über das Internet bereits verfügbar sind. Die Umsetzung einer durchgängigen Infrastruktur ist daher in sehr viel größerem Ausmaß von sozialen und ökonomischen Fragestellungen abhängig.

Die große Zahl an Forschungsprojekten im Bereich digitaler Bibliotheken und die Aktivitäten der wissenschaftlichen Fachverlage spiegeln das große Interesse an elektronisch verfügbaren Artikeln von seiten der *scientific community* wider. Die Akzeptanz elektronischer Fachzeitschriften und -artikel ist zwar in hohem Maße von der Fachdisziplin abhängig, dennoch ist die Akzeptanz über alle akademischen Disziplinen hinweg wohl eher eine Frage der Zeit.

Die möglichen ökonomischen Implikationen einer Transformation des wissenschaftlichen Kommunikationssystems von gedruckten Artikeln in Fachzeitschriften zu elektronischen Aufsätzen wurden in dieser Arbeit explizit ausgeklammert. Dennoch bietet dieses Thema einigen Diskussionsbedarf, gerade im Hinblick auf den Erfolg von kostenfreien Preprint-Servern und Archiven. Im wesentlichen existieren hier zwei Extrempositionen: Auf der einen Seite die Verlage, die die Rollenverteilung im traditionellen Publikationssystem und ihre Profite in die elektronische Welt hinüberretten

möchten, und auf der anderen Seite eine wachsende Zahl von Wissenschaftlern, die sich in verschiedenen Projekten für eine kostenfreie Publikation von wissenschaftlichen Artikeln einsetzen (vgl. Okerson und O'Donnel, 1995).

Welche Geschäftsmodelle sich durchsetzen werden, ist nicht abzusehen; mit Sicherheit sagen lässt sich allerdings, dass die Zukunft der wissenschaftlichen Kommunikation im Internet liegt.

Literaturverzeichnis

- [Apps und MacIntyre 2000] APPS, Ann ; MACINTYRE, Ross: **Dublin Core Metadata for Electronic Journals**. In: *ECDL2000 - Fourth European Conference on Research and Advanced Technology for Digital Libraries, Lisbon, Portugal, 18-20 September 2000*, URL <http://epub.mimas.ac.uk/papers/appsmacecdl2000.pdf>, 2000 (To be published: Lecture Notes in Computer Science, Springer-Verlag)
- [Arms 1999a] ARMS, Caroline (Hrsg.): **Enabling Access in Digital Libraries: A Report on a Workshop on Access Management**. The Digital Library Federation, 1999
- [Arms 1995] ARMS, William Y.: **Key Concepts in the Architecture of the Digital Library**. In: *D-Lib Magazine* 1 (1995), Nr. 7. – URL <http://www.dlib.org/dlib/July95/07arms.html>. – Zugriffsdatum: 28.10.2000
- [Arms 1999b] ARMS, William Y.: **Preservation of Scientific Serials: Three Current Examples**. In: *Journal of Electronic Publishing* 5 (1999), Nr. 2. – URL <http://www.press.umich.edu/jep/05-02/arms.html>. – Zugriffsdatum: 28.10.2000
- [Arms 2000] ARMS, William Y.: **Digital Libraries**. Cambridge, Massachusetts : MIT Press, 2000 (Digital Libraries and Electronic Publishing). – ISBN 0-262-01880-8
- [Arms u. a. 1996] ARMS, William Y. ; DAIGLE, Leslie ; DANIEL, Ron ; LALIBERTE DAN MEALLING, Michael ; MOORE, Keith ; WEIBEL, Stuart: **Unifom Resource Names - A Progress Report**. In: *D-Lib Magazine* 2 (1996), Nr. 2. – URL <http://www.dlib.org/dlib/July95/07arms.html>. – Zugriffsdatum: 28.10.2000
- [Arps 1993] ARPS, Mark: **CD-ROM: Archival Considerations**. In: *Preservation of Electronic Formats & Electronic Formats For Preservation*. Highsmith Press, 1993

- [Atkins 1999] ATKINS, Helen: **The ISI Web of Science - Links and Electronic Journals**. In: *D-Lib Magazine* 9 (1999), Nr. 5. – URL <http://www.dlib.org/dlib/september99/atkins/09atkins.html>. – Zugriffsdatum: 28.10.2000
- [Atkins u. a. 2000] ATKINS, Helen ; LYONS, Catherine ; RATNER, Howard ; RISHER, Carol ; SHILLUM, Chris ; SIDMAN, David ; STEVENS, Andres: **Reference Linking with DOIs: A Case Study**. In: *D-Lib Magazine* 6 (2000), Nr. 2. – URL <http://www.dlib.org/dlib/february00/02risher.html>. – Zugriffsdatum: 28.10.2000
- [Baker 1998] BAKER, Thomas: **Languages for Dublin Core**. In: *D-Lib Magazine* 4 (1998), Nr. 12. – URL <http://www.dlib.org/dlib/december98/12baker.html>. – Zugriffsdatum: 28.10.2000
- [Barth 1996] BARTH, R. **Archive/Bibliotheken in Mesopotamien und Ägypten**. Dezember 1996. – URL <http://biblio.unibe.ch/stub/vor196/03/mes.html>. – Zugriffsdatum: 13.11.2000
- [Bearman u. a. 1999] BEARMAN, David ; RUST, Godfrey ; WEIBEL, Stuart I. ; MILLER, Eric ; TRANT, Jennifer: **A Common Model to Support Interoperable Metadata** - Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. In: *D-Lib Magazine* 5 (1999), Nr. 1. – URL <http://www.dlib.org/dlib/january99/bearman/01bearman.html>. – Zugriffsdatum: 28.10.2000
- [Bearman und Sochats 1994] BEARMAN, David ; SOCHATS, Ken. **Metadata Requirements for Evidence**. 1994. – URL <http://www.lis.pitt.edu/~nhprc/BACartic.html>. – Zugriffsdatum: 25.07.2000
- [Berners-Lee 1998] BERNERS-LEE, Tim: **Web Architecture from 50,000 feet**. 1998. – URL <http://www.w3.org/DesignIssues/Architecture.html>. – Zugriffsdatum: 30.10.2000. – *Id*: *Architecture.html,v1.562000/09/0817:38:37timblExp*
- [Bide 1997] BIDE, Mark. **In Search of the Unicorn - The Digital Object Identifier from a User Perspective**. BNBRF Report 89 (British Library Research & Innovation Report 84). 1997. – URL <http://www.bic.org.uk/unicorn2.pdf>. – Zugriffsdatum: 14.07.2000

- [Bide und Green 1997] BIDE, Mark ; GREEN, Brian. **Unique Identifiers: a brief introduction.** 1997. – URL <http://www.bic.org.uk/uniqid.html>. – Zugriffsdatum: 21.09.2000
- [Böhle 1997] BÖHLE, Knud: **Elektronisches Publizieren.** In: BUDER, Marianne (Hrsg.) ; REHFELD, Werner (Hrsg.) ; SEEGER, Thomas (Hrsg.) ; STRAUCH, Dietmar (Hrsg.): *Grundlagen der praktischen Information und Dokumentation.* 4. Auflage. München; New Providence; London; Paris : KG Saur, 1997, S. 397–424
- [Borggraefe 1999] BORGGRAEFE, Stefan: **Reiche Ernte - Hierarchische Suchindizes mit Harvest.** In: *iX* (1999), Nr. 3, S. 126–129
- [Butler 1999] BUTLER, Declan: **The Writing is on the Web for Science Journals in Print.** In: *Nature* 21 (1999), Nr. Januar 1999, S. 195–200. – URL http://cas.bellarmine.edu/tietjen/images/writing_is_on_the_web_for_scienc.htm. – Zugriffsdatum: 28.06.2000
- [Caplan 1997] CAPLAN, Priscilla: **To Hel(sinki) and Back for the Dublin Core.** In: *Public-Access Computer Systems Review* 8 (1997), Nr. 4
- [Caplan 1999] CAPLAN, Priscilla: **A model for reference linking.** 10 Jun 1999 1999. – URL <http://www.lib.uchicago.edu/Annex/pcaplan/reflink.html>. – Zugriffsdatum: 28.10.2000. – Report of the working group of the reference linking workshop (NISO)
- [Caplan und Arms 1999] CAPLAN, Priscilla ; ARMS, William Y.: **Reference Linking for Journal Articles.** In: *D-Lib Magazine* 5 (1999), Nr. 7. – URL <http://www.dlib.org/dlib/july99/caplan/07caplan.html>. – Zugriffsdatum: 28.10.2000
- [Carr u. a. 1995] CARR, L. ; DE ROURE, D. ; HALL, W. ; HILL, G.: **The Distributed Link Service: a Tool for Publishers, Authors and Readers.** In: *World Wide Web Journal (special issue, Proceedings of the Fourth International World Wide Web Conference)* (1995), Nr. 1. – URL <http://www.w3.org/pub/Conferences/WWW4/Papers/178/>. – Zugriffsdatum: 04.10.2000
- [Causton 1998] CAUSTON, Laurie: **Identifying and Describing Web Resources / European Commission DGXIII/E-4.** URL http://www.elpub.org/assets/images/web_resources.pdf. – Zugriffsdatum: 28.10.2000, 1998. – Forschungsbericht

- [Cox u. a. 2000] COX, Simon ; MILLER, Eric ; POWELL, Andy. **Recording Qualified Dublin Core metadata in HTML meta elements.** 15.08. 2000. – URL <http://purl.org/dc/documents/wd/dcq-html-20000815.htm>. – Zugriffsdatum: 28.08.2000
- [Cuenca u. a. 1999] CUENCA, Pedro ; SOSA, Vicente ; ROMERO, Julian ; HERNANZ, Israel: **Lessons Learned from the Early Adoption of URNs in an Intranet Environment.** In: *INET'99 Proceedings*, URL http://www.isoc.org/inet99/proceedings/4m/4m_2.htm. – Zugriffsdatum: 23.10.2000, 1999
- [Dalitz u. a. 1997] DALITZ, Wolfgang ; GRÖTSCHER, Martin ; LÜGGER, Joachim: **Information Services for Mathematics in the Internet (Math-Net).** In: SYDOW, A. (Hrsg.): *Proc. of the 15th IMACS World Congress on Scientific Computation: Modelling and Applied Mathematics*, Wissenschaft und Technik Verlag, 1997, S. 773–778. – URL <ftp://ftp.zib.de/pub/zib-publications/reports/SC-97-25.ps.Z>. – Zugriffsdatum: 28.10.2000
- [Daly 1998] DALY, Bernadette: **Launching an Electronic Magazine: an overview of value-added features and services.** In: *Ariadne* 1998 (1998), Nr. 18. – URL <http://www.ariadne.ac.uk/issue18/web-magazine/intro.html>. – Zugriffsdatum: 28.10.2000
- [Daniel und Mealling 1997] DANIEL, Ronald ; MEALLING, M. **Resolution of Uniform Resource Identifiers using the Domain Name System (RFC 2168).** Juni 1997. – URL <http://www.ietf.org/rfc/rfc2168.txt>. – Zugriffsdatum: 13.09.2000
- [Davidson und Douglas 1998] DAVIDSON, Lloyd ; DOUGLAS, Kimberly: **Digital Object Identifier: Promise and Problems for Scholarly Publishing.** In: *Journal of Electronic Publishing* 4 (1998), Nr. 2. – URL <http://www.press.umich.edu/jep/04-02/davidson.html>. – Zugriffsdatum: 28.10.2000
- [DC-FAQ 2000] **Dublin Core Metadata Initiative / Question and Answers.** 28. März 2000. – URL <http://purl.org/DC/education/index.htm>. – Zugriffsdatum: 24.08.2000
- [DCMES-QUALIFIERS 2000] **Dublin Core Qualifiers.** 11. Juli 2000. – URL <http://purl.org/dc/documents/dcmes-qualifiers>. – Zugriffsdatum: 24.08.2000

- [Dempsey und Heery 1997] DEMPSEY, Lorcan ; HEERY, Rachel. **A review of metadata: a survey of current resource description formats.** März 1997. – URL <http://www.ukoln.ac.uk/metadata/desire/overview/>. – Zugriffsdatum: 31.07.2000
- [Dempsey und Weibel 1996] DEMPSEY, Lorcan ; WEIBEL, Stuart: **The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description.** In: *D-Lib Magazine* 2 (1996), Nr. 7. – URL <http://www.oclc.org/oclc/research/publications/review96/warwick.htm>. – Zugriffsdatum: 15.08.2000
- [Diedrich 1999] DIEDRICH, Oliver: **Nie mehr das Paßwort vergessen - Biometrische Systeme zur Identifikation und Authentifizierung.** In: *c't* (1999), Nr. 6, S. 188–191
- [Doyle 2000] DOYLE, Mark: **Pragmatic citing and linking in electronic scholarly publishing.** In: *Learned Publishing* 13 (2000), Januar, Nr. 1, S. 5–14
- [Endres und Fellner 2000] ENDRES, Albert ; FELLNER, Dieter W.: **Digitale Bibliotheken: Informatik-Lösungen für globale Wissensmärkte.** Heidelberg : dpunkt-Verlag, 2000
- [Felt u. a. 1995] FELT, Ulrike ; NOWOTNY, Helga ; TASCHWER, Klaus: **Wissenschaftsforschung: eine Einführung.** Bd. 1086: Studium. Reihe Campus. Frankfurt/Main : Campus Verlag, 1995
- [Fisher 1996] FISHER, J.: **Traditional Publishers and Electronic Journals.** In: PEEK, Robin P. (Hrsg.) ; NEWBY, Gregory B. (Hrsg.): *Scholarly Publishing: The Electronic Frontier.* Cambridge, MA : MIT Press, 1996
- [FRBR 1998] IFLA Study Group on the Functional Requirements for Bibliographic Records (Hrsg.): **Functional requirements for bibliographic records : final report.** 1998. – URL <http://www.ifla.org/VII/s13/frbr/frbr.pdf>. – Zugriffsdatum: 11.11.2000. – Saur; München; 3-598-11382-X
- [Garfield 1955] GARFIELD, Eugene: **Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas.** In: *Science* 122 (1955), Nr. 3159, S. 108–111. – URL [http://www.garfield.library.upenn.edu/papers/science_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html). – Zugriffsdatum: 28.10.2000

- [Garfield 1977] GARFIELD, Eugene: **Negative Science and "The outlook for the flying machine"**. In: *Essays of an Information Scientist* 3 (1977), S. 155–172
- [Garfield 1986] GARFIELD, Eugene: **Refereeing and Peer Review. Part 1. Opinion and Conjecture on the Effectiveness of Refereeing**. In: *Current Contents* (1986), Nr. 31, S. 3–11
- [Garfield 1994] GARFIELD, Eugene. **The Impact Factor**. 1994. – URL <http://www.isinet.com/isi/hot/essays/7>. – Zugriffsdatum: 05.11.2000
- [Garfield 1996] GARFIELD, Eugene: **What is the Primordial Reference for the Phrase "Publish or Perish"?** In: *The Scientist* 10 (1996), Nr. 12, S. 11. – URL http://www.the-scientist.com/yr1996/june/comm_960610.html. – Zugriffsdatum: 28.10.2000
- [Garvey 1979] GARVEY, William D.: **Communication: The Essence of Science**. Pergamon Press, 1979
- [Garvey und Griffith 1979] GARVEY, William D. ; GRIFFITH, Belver C.: **Communication and Information Processing within Scientific Disciplines: Empirical Findings for Psychology**. Siehe (Garvey, 1979), Kap. Appendix A
- [Gerck 2000] GERCK, Ed. **Overview of Certification Systems: X.509, PKIX, CA, PGP & SKIP**. 2000. – URL <http://www.mcg.org.br/certover.pdf>. – Zugriffsdatum: 12.10.2000
- [Ginsparg 1994] GINSPARG, Paul: **First Steps towards Electronic Research Communication**. In: *Computers in Physics* 8 (1994), Nr. 4, S. 390–396. – URL <http://xxx.lanl.gov/blurb/>
- [Grötschel und Lügger 1996] GRÖTSCHEL, M. ; LÜGGER, J.: **Wissenschaftliche Information und Kommunikation im Umbruch - über das Internet zu neuen wissenschaftlichen Informationsinfrastrukturen**. In: *Forschung & Lehre* (1996), Nr. 1, S. 194–198. – URL http://elib.zib.de:888/0x82496a1f_0x000e0fb7. – Zugriffsdatum: 30.06.2000
- [Hakala u. a. 1998] HAKALA, Juha ; HANSEN, Preben ; HUSBY, Ole ; KOCH, Traugott ; THORBORG, Susanne: **The Nordic Metadata Project - Final Report** / Helsinki University Library. URL <http://linnea.helsinki.fi/meta/nmfinal.doc>. – Zugriffsdatum: 24.08.2000, 1998. – Forschungsbericht. – ISBN 951-45-8247-0

- [Harmsz u. a. 2000] HARMSZ, F.A.P. ; TOL, M.C. van d. ; KIRCZ, J. **A modular structure for electronic scientific articles.** 2000. – URL <http://www.science.uva.nl/projects/commphys/papers/infwet/infwet.html>. – Zugriffsdatum: 04.08.2000
- [Harnad 1991] HARNAD, Stevan: **Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge.** In: *Public-Access Computer Systems Review* 2 (1991), Nr. 1. – URL <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html>. – Zugriffsdatum: 30.10.2000
- [Harnad 1996] HARNAD, Stevan: **Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals.** In: PEEK, Robin P. (Hrsg.) ; NEWBY, Gregory B. (Hrsg.): *Scholarly Publishing - The Electronic Frontier.* 1996
- [Heery u. a. 2000] HEERY, Rachel ; GARDNER, Tracy ; DAY, Michael ; PATEL, Manjula: **DESIRE metadata registry framework** / UKOLN: the UK Office for Library and Information Networking. URL <http://www.desire.org/html/research/deliverables/D3.5/>. – Zugriffsdatum: 07.09.2000, 2000. – Forschungsbericht
- [Hellman 1999a] HELLMAN, Eric. **Scholarly Link Specification Framework.** November 1999. – URL <http://www.openly.com/slinks/SLinkS.html>. – Zugriffsdatum: 06.09.2000
- [Hellman 1999b] HELLMAN, Eric: **The S-Link-S Framework for Reference Linking: Architecture and Implementation.** In: SMITH, John W. (Hrsg.) ; ARDÖ, Anders (Hrsg.) ; RONNEBY, Peter L. (Hrsg.): *Proceedings of the ICCC/IFIP Conference on Electronic Publishing '99. Redefining the Information Chain - New Ways and Voices,* ICCC Press, 1999
- [Hitchcock 1998] HITCHCOCK, Steve. **Open Journal project: final report to eLib.** November 1998. – URL <http://journals.ecs.soton.ac.uk/yr3/3rd-year-open.htm>. – Zugriffsdatum: 28.10.2000
- [Hitchcock u. a. 1998] HITCHCOCK, Steve ; CARR, Leslie ; HALL, Wendy ; HARRIS, Steve: **Linking Electronic Journals: Lessons from the Open Journal Project.** In: *D-Lib Magazine* 4 (1998), Nr. December. – URL <http://www.dlib.org/dlib/december98/12hitchcock.html>. – Zugriffsdatum: 28.10.2000

- [Hunter 1998] HUNTER, Karen: **Adding Value by Adding Links**. In: *Journal of Electronic Publishing* 3 (1998), Nr. 3. – URL <http://www.press.umich.edu/jep/03-03/hunter.html>. – Zugriffsdatum: 28.10.2000
- [Hurd 1996] HURD, Julie M.: **Models of Scientific Communication Systems**. In: CRAWFORD, Susan Y. (Hrsg.) ; HURD, Julie M. (Hrsg.) ; WALKER, A.C. (Hrsg.): *From Print to Electronic: The Transformation of Scientific Communication*. Medford, NJ : Information Today Inc., 1996, S. 9–33
- [Ianella u. a. 1996] IANELLA, Renato ; SUE, Hoyleen ; LEONG, Danny. **BURNS: Basic URN Service resolution for the Internet**. 1996. – URL <http://www.dstc.edu.au/RDU/publications/apweb96/>. – Zugriffsdatum: 13.11.2000
- [Kengeri u. a. 1999] KENGERI, Rekha ; SEALS, Cheryl D. ; HARLEY, Hope D. ; REDDY, Himabindu P. ; FOX, Edward A.: **Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD**. In: *International Journal on Digital Libraries* 2 (1999), S. 157–169
- [King und Tenopir 1998] KING, Donald W. ; TENOPIR, Carol. **Economics Cost Models of Scientific Scholarly Journals**. 1998. – URL <http://www.bodley.ox.ac.uk/icsu/kingppr.htm>. – Zugriffsdatum: 22.07.2000
- [Kling und Covi 1997] KLING, Rob ; COVI, Lisa. **Digital Libraries and the Practices of Scholarly Communication**. 7 Januar 1997. – URL <http://www-slis.lib.indiana.edu/kling/Scit/SCIT97.HTM>. – Zugriffsdatum: 28.10.2000
- [Kling und McKim 1999] KLING, Rob ; MCKIM, Geoffrey: **Scholarly Communication and the Continuum of Electronic Publishing**. In: *Journal of the American Society for Information Science* 50 (1999), Nr. 10, S. 890–906
- [Krieb 1999] KRIEB, Dennis: **You Can't Get There from Here: Issues in Remote Access to Electronic Journals for a Health Sciences Library**. In: *Issues in Science and Technology Librarianship* Spring 99 (1999). – URL <http://www.library.ucsb.edu/istl/99-spring/article3.html>. – Zugriffsdatum: 28.10.2000
- [Lagoze 2000] LAGOZE, Carl: **Accommodating Simplicity and Complexity in Metadata: Lessons from the Dublin Core Experience** / Cornell University. URL <http://www.ncstrl.org/Dienst/UI/1.0/Display/>

- ncstrl.cornell/TR2000-1801. – Zugriffsdatum: 10.08.2000, 2000. – Forschungsbericht
- [Lagoze u. a. 1996] LAGOZE, Carl ; LYNCH, Clifford ; DANIEL, Ron J.: **The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata** / Cornell University. URL <http://cs-tr.cs.cornell.edu:80/Dienst/Repository/2.0/Body/ncstrl.cornell/TR96-1593/>. – Zugriffsdatum: 14.07.2000, Juni 1996 (TR 96-1593). – Forschungsbericht
- [Lassila 1998] LASSILA, Ora: **Web Metadata: A Matter of Semantics**. In: *IEEE Internet Computing* (1998), July/August 1998, S. 30–37
- [Lawrence u. a. 1999a] LAWRENCE, Steve ; BOLLACKER, Kurt ; GILES, C. L.: **Digital Libraries and Autonomous Citation Indexing**. In: *IEEE Computer* 32 (1999), Nr. 6, S. 67–71
- [Lawrence u. a. 1999b] LAWRENCE, Steve ; BOLLACKER, Kurt ; GILES, C. L.: **Indexing and Retrieval of Scientific Literature**. In: *Eighth International Conference on Information and Knowledge Management, CIKM 99*. Kansas City, Missouri, November 1999, S. 139–146
- [Lederberg 1993] LEDERBERG, Joshua: **Communication as the root of scientific progress**. In: *The Scientist* 10 (1993), Nr. 3. – URL http://www.the-scientist.com/yr1993/feb/opin_930208.html
- [Li und Crane 1996] LI, Xia ; CRANE, Nancy: **Bibliographic Formats for Citing Electronic Information. The Official Internet World Guide to Electronic Styles: A Handbook to Citing Electronic Information**. Mecklermedia, 1996
- [Lynch und Garcia-Molina 1995] LYNCH, Clifford ; GARCIA-MOLINA, Hector. **Interoperability, Scaling, and the Digital Libraries Research Agenda: A Report on the May 18-19, 1995 IITA Digital Libraries Workshop**. 22. August 1995. – URL <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>. – Zugriffsdatum: 15.07.2000
- [Lynch u. a. 1998] LYNCH, Clifford ; PRESTON, C. ; DANIEL, Ron. **Using Existing Bibliographic Identifiers as Uniform Resource Names (RFC 2288)**. Februar 1998. – URL <http://www.ietf.org/rfc/rfc2288.txt>. – Zugriffsdatum: 13.09.2000

- [Lynch 1996] LYNCH, Clifford A.: **Integrity Issues in Electronic Publishing**. In: PEEK, Robin P. (Hrsg.) ; NEWBY, Gregory B. (Hrsg.): *Scholarly Publishing: the electronic frontier*. Cambridge, Massachusetts : The MIT Press, 1996, Kap. 8, S. 133–145. – ISBN 0–262–16157–5
- [Lynch 1997] LYNCH, Clifford A.: **Searching the Internet**. In: *Scientific American* (1997), Nr. 3. – URL <http://www.sciam.com/0397issue/0397lynch.html>. – Zugriffsdatum: 15.07.2000
- [Lynch 1998] LYNCH, Clifford A.: **Access Management for Networked Information Resources**. In: *ARL: A Bimonthly Report on Research Libraries Issues and Actions from ARL, CNI, and SPARC* (1998), Nr. 201. – URL <http://www.arl.org/newsltr/201/cni.html>. – Zugriffsdatum: 28.10.2000
- [Lyons und Ratner 1999] LYONS, Catherine ; RATNER, Howard. **DOIs Used for Reference Linking: DOI-X**. 1999. – URL <http://dx.doi.org/10.1000/6382-1>
- [Martin 1998] MARTIN, David: **Beyond Dublin Core - the need for high quality product information**. 1998. – URL <http://www.bic.org.uk/beyonddc.rtf>. – Zugriffsdatum: 05.11.2000. – Paper presented at BIC / BL Seminar Trading Electronic Content, 10 March 1998
- [Mason und Sutton 2000] MASON, Jon ; SUTTON, Stuart. **Education Working Group: Draft Proposal**. 30.04. 2000. – URL <http://purl.org/dc/documents/education-wd-20000430.htm>. – Zugriffsdatum: 28.08.2000
- [Math-Net 1999] GRÖTSCHHEL, Martin (Hrsg.) ; DALITZ, Wolfgang (Hrsg.): **Math-Net Abschlussbericht** / Konrad-Zuse-Zentrum für Informationstechnik, Berlin. URL <ftp://ftp.zib.de/pub/zib-publications/reports/SC-99-24.ps>. – Zugriffsdatum: 10.08.2000, August 1999 (SC 99-24). – Forschungsbericht
- [Merton 1968] MERTON, Robert K.: **The Matthew Effect in Science**. In: *Science* 159 (1968), Nr. 3810, S. 56–63
- [Merton 1988] MERTON, Robert K.: **The Matthew Effect in Science, II Cumulative Advantage and the Symbolism of Intellectual Property**. In: *ISIS* 79 (1988), S. 606–623

- [MetaForm 1997–2000] **SUB Göttingen MetaForm**. 1997-2000. – URL <http://www2.sub.uni-goettingen.de/metaform/index.html>. – Zugriffsdatum: 12.09.2000
- [Miller 1998] MILLER, Eric: **An Introduction to the Resource Description Framework**. In: *D-Lib Magazine* 4 (1998), Nr. 5. – URL <http://www.dlib.org/dlib/may98/miller/05miller.html>. – Zugriffsdatum: 28.10.2000
- [Miller 1999] MILLER, Paul: **Z39.50 for All**. In: *Ariadne* (1999), Nr. 21. – URL <http://www.ariadne.ac.uk/issue21/z3950/intro.html>. – Zugriffsdatum: 28.10.2000
- [Miller 2000a] MILLER, Paul: **I am a name and a number**. In: *Ariadne* (2000), Nr. 24. – URL <http://www.ariadne.ac.uk/issue24/metadata/>. – Zugriffsdatum: 28.10.2000
- [Miller 2000b] MILLER, Paul: **Interoperability - What is it and Why should I want it?** In: *Ariadne* (2000), Nr. 24. – URL <http://www.ariadne.ac.uk/issue24/interoperability/>. – Zugriffsdatum: 28.10.2000
- [Moats 1997] MOATS, R. **URN Syntax (RFC 2141)**. Mai 1997. – URL <http://www.ietf.org/rfc/rfc2141.txt>. – Zugriffsdatum: 13.09.2000
- [Morgan 1999] MORGAN, Cliff. **DC Bibliographic Citation Working Draft (02.07.1999)**. 1999. – URL <http://purl.org/dc/documents/wd-citation-19990702.htm>. – Zugriffsdatum: 03.11.2000
- [Morton 1997] MORTON, Bruce: **Is the Journal as We Know it an Article of Faith? An Open Letter to the Faculty**. In: *Public-Access Computer Systems Review* 8 (1997), Nr. 2
- [Nat 1996] National Information Standards Organization (U.S.): **Serial item and contribution identifier (SICI)**. 1996. – developed by the National Information Standards Organization : approved August 14, 1996 by the American National Standards Institute
- [Ockenfeld 1997] OCKENFELD, Marlies: **Klassische Informationsdienste**. In: BUDER, Marianne (Hrsg.) ; REHFELD, Werner (Hrsg.) ; SEEGER, Thomas (Hrsg.) ; STRAUCH, Dietmar (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. 4. Auflage. München : KG Saur, 1997, S. 257–279

- [Odlyzko 1996] ODLYZKO, Andrew M.: **Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals.** In: PEEK, Robin P. (Hrsg.) ; NEWBY, Gregory B. (Hrsg.): *Scholarly Publishing: The Electronic Frontier.* Cambridge, MA : MIT Press, 1996, S. 91–102
- [Odlyzko 1997] ODLYZKO, Andrew M.: **The Economics of Electronic Journals.** In: *Scholarly Communication and Technology.* Emory University, 24.05.1997-25.05.1997 1997. – URL <http://www.arl.org/scomm/scat/odlyzko.html>
- [Okerson 1991] OKERSON, Ann: **The Electronic Journal: What, Whence, and When?** In: *The Public-Access Computer Systems Review* 2 (1991), Nr. 1
- [Okerson und O'Donnel 1995] OKERSON, Ann S. ; O'DONNELL, James J. **Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing.** 1995. – URL <http://www.arl.org/scomm/subversive/toc.html>. – Zugriffsdatum: 02.08.2000
- [Paskin 1997] PASKIN, Norman: **Information Identifiers.** In: *Learned Publishing* 10 (1997), Nr. 2, S. 135–156. – URL <http://www.elsevier.nl/homepage/about/infoident/>. – Zugriffsdatum: 21.12.2000
- [Paskin 1999] PASKIN, Norman: **DOI: Current Status and Outlook.** In: *D-Lib Magazine* 5 (1999), Nr. 5. – URL <http://www.dlib.org/dlib/may99/05paskin.html>. – Zugriffsdatum: 28.10.2000
- [Paskin 2000] PASKIN, Norman. **The DOI Handbook (Version 0.3).** Juli 2000. – URL <http://dx.doi.org/10.1000/182>
- [Pommerening 1993] POMMERENING, Klaus: **Krebsregister - Ein Konzept für die sozialverträgliche Gestaltung.** In: *F!FF Kommunikation* (1993), Nr. 4, S. 64–66. – URL <http://www.uni-mainz.de/~pommeren/Artikel/krebsreg.html>. – Zugriffsdatum: 17.08.2000
- [Rambler 1999] RAMBLER, Mark: **Do it yourself? A New Solution to the Journal Crisis.** In: *Journal of Electronic Publishing* 4 (1999), Nr. 3. – URL <http://www.press.umich.edu/jep/04-03/rambler.html>. – Zugriffsdatum: 28.10.2000
- [rdf-schema 2000] BRICKLEY, Dan (Hrsg.) ; GUHA, R.V. (Hrsg.): **Resource Description Framework (RDF) Schema Specification 1.0.** 27 März 2000. – URL <http://www.w3.org/2000/01/rdf-schema/>

- [//www.w3.org/TR/rdf-schema](http://www.w3.org/TR/rdf-schema). – Zugriffsdatum: 11.11.2000. – W3C Candidate Recommendation 27 March 2000
- [REC-DCES 1999] **Dublin Core Metadata Element Set, Version 1.1: Reference Description**. 2. Juli 1999. – URL <http://purl.org/dc/documents/rec-dces-19990702.htm>. – Zugriffsdatum: 24.08.2000
- [REC-rdf-syntax 1999] LASSILA, Ora (Hrsg.) ; SWICK, Ralph R. (Hrsg.): **Resource Description Framework (RDF) Model and Syntax Specification**. 22 Februar 1999. – URL <http://www.w3.org/TR/REC-rdf-syntax/>. – Zugriffsdatum: 11.11.2000. – W3C Recommendation
- [van Rooyen 1998] ROOYEN, Susan van: **A critical examination of the peer review process**. In: *Learned Publishing* (1998), Nr. 11, S. 185–191
- [Rosenblatt 1997] ROSENBLATT, Bill: **Solving the Dilemma of Copyright Protection Online**. In: *Journal of Electronic Publishing* 3 (1997), Dezember, Nr. 2. – URL <http://www.press.umich.edu/jep/03-03/doi.html>. – Zugriffsdatum: 15.09.2000
- [Rusch-Feja 1999] RUSCH-FEJA, Diann: **Digital Libraries - Informationsform der Zukunft für die Informationsversorgung und Informationsbereitstellung? (Teil 2)**. In: *B.I.T. Online* (1999), Nr. 3, S. 281–306
- [Rusch-Feja und Becker 1999] RUSCH-FEJA, Diann ; BECKER, Hans J.: **Global Info - The German Digital Libraries Project**. In: *D-Lib Magazine* 5 (1999), Nr. 4. – URL <http://www.dlib.org/dlib/april99/04rusch-feja.html>. – Zugriffsdatum: 28.10.2000
- [Rust und Bide] RUST, Godfrey ; BIDE, Marc. **indecs metadata schema: building blocks**. Version published at indecs conference Names, Numbers and Networks, Washington DC, November 15, 1999
- [Seglen 1997] SEGLEN, Per O.: **Why the impact factor of journals should not be used for evaluating research**. In: *British Medical Journal* 314 (1997), S. 497. – URL <http://www.bmj.com/cgi/content/full/314/7079/497>. – Zugriffsdatum: 01.10.2000
- [Sietmann 1999] SIETMANN, Richard: **Zirkelspiele - Die wissenschaftliche Literaturversorgung steckt in der Krise**. In: *c't* (1999), Nr. 20, S. 216–231

- [Sietmann 2000a] SIETMANN, Richard. **Eine Zeitschrift auf den Server einer Universität zu legen, ist keine große Leistung** *Ein Gespräch mit Ar-noud deKemp vom wissenschaftlichen Springer-Verlag.* 12. Januar 2000. – URL <http://www.heise.de/tp/deutsch/html/result.xhtml?url=/tp/deutsch/inhalt/on/5673/1.html&words=Sietmann>. – Zugriffsdatum: 28.10.2000
- [Sietmann 2000b] SIETMANN, Richard: **Online shoppen, offline bezahlen - Das Cybergeld kommt nicht aus den Startlöchern.** In: *c't* (2000), Nr. 9, S. 44
- [Smith 1997] SMITH, John W. T.: **The Deconstructed Journal.** 30 Sep 1998 1997. – URL <http://library.ukc.ac.uk/library/ICCC/papers/deconjnl.htm>. – Published in Electronic Publishing '97 - New Models and Opportunities - Proceedings of the ICC/IFIP Conference on Electronic Publishing, 14-16 April 1997, University of Kent at Canterbury, UK, Edited by: Fytton Rowland and Jack Meadows ICC Press 1997,
- [Solbakk] SOLBAKK, Svein A. **Long term preservation of electronic material : preliminary experiences from the Norwegian National Library.** – URL <http://www.nb.no/rapporter/preserv.html>. – Zugriffsdatum: 27.08.2000
- [de Solla Price 1986] SOLLA PRICE, Derek J.: **Little Science, Big Science ... and beyond.** Columbia University Press, 1986
- [Sollins 1998] SOLLINS, Karen. **Architectural Principles of Uniform Resource Name Resolution (RFC 2276).** Januar 1998. – URL <http://www.ietf.org/rfc/rfc2276.txt>. – Zugriffsdatum: 13.09.2000
- [Sollins und Masinter 1994] SOLLINS, Karen ; MASINTER, L. **Functional Requirements for Uniform Resource Names (RFC 1737).** Dezember 1994. – URL <http://www.ietf.org/rfc/rfc1737.txt>. – Zugriffsdatum: 13.09.2000
- [Sommerlad 1998] SOMMERLAD, Elizabeth. **The Archiving of Electronic Scholarly Information: An overview.** 1998. – URL <http://www.unison.nsw.edu.au/esomer.html>. – Zugriffsdatum: 16.08.2000
- [Van de Sompel und Hochstenbach 1999a] SOMPEL, Herbert Van d. ; HOCHSTENBACH, Patrick: **Reference Linking in a Hybrid Library Environment - Part 1: Frameworks for Linking.** In: *D-Lib Magazine* 5 (1999), Nr. 4.

- URL http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html. – Zugriffsdatum: 11.09.2000
- [Van de Sompel und Hochstenbach 1999b] SOMPEL, Herbert Van d. ; HOCHSTENBACH, Patrick: **Reference Linking in a Hybrid Library Environment - Part 2: SFX, a Generic Linking Solution**. In: *D-Lib Magazine* 5 (1999), Nr. 4. – URL http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html. – Zugriffsdatum: 11.09.2000
- [Van de Sompel und Hochstenbach 1999c] SOMPEL, Herbert Van d. ; HOCHSTENBACH, Patrick: **Reference Linking in a Hybrid Library Environment - Part 3: Generalizing the SFX solution in the "SFX@Ghent & SFX@LANL" experiment**. In: *D-Lib Magazine* 5 (1999), Nr. 10. – URL http://www.dlib.org/dlib/october99/van_de_sompel/10van_de_sompel.html. – Zugriffsdatum: 11.09.2000
- [Stegmann 1999] STEGMANN, Johannes: **Why not use Journal Impact Factors?** In: *EAHIL Newsletter* (1999), Mai. – URL <http://www.eahil.org/newsletter/47/impact.htm>. – Zugriffsdatum: 01.10.2000
- [Sun und Lannom 2000] SUN, Sam X. ; LANNOM, Larry. **Handle System Overview**. August 2000. – URL <http://www.ietf.org/internet-drafts/draft-sun-handle-system-05.txt>. – Zugriffsdatum: 13.09.2000
- [Taubes 1996] TAUBES, Gary: **Science Journals Go Wired**. In: *Science* 271 (1996), Nr. 5250, S. 764. – URL <http://www.epub.org.br/papers/sciwirl1.htm>. – Zugriffsdatum: 22.07.2000
- [Varian 1997] VARIAN, Hal R.: **The Future of Electronic Journals**. In: *Scholarly Communication and Technology*. Emory University, 24.05.1997-25.05.1997 1997. – URL <http://www.arl.org/scomm/scat/varian.html>. – Zugriffsdatum: 17.08.2000
- [Wasmeier 2000] WASMEIER, Michael: **Web-Währungen - Online-Bezahlungsverfahren für ECommerce**. In: *c't* (2000), Nr. 11, S. 152–157
- [Weibel 1995] WEIBEL, Stuart I.: **Metadata: The Foundations of Resource Description**. In: *D-Lib Magazine* 1 (1995), Nr. 7. – URL <http://www.dlib.org/dlib/July95/07weibel.html>. – Zugriffsdatum: 28.10.2000

- [Weibel 1997] WEIBEL, Stuart I.: **The Evolving Metadata Architecture for the World Wide Web: Bringing Together the Semantics, Structure and Syntax of Resource Description**. In: *International Symposium on Research, Development and Practice in Digital Libraries : ISDL'97*, URL <http://www.DL.ulis.ac.jp/ISDL97/proceedings/weibe.html>. – Zugriffsdatum: 28.10.2000, 1997
- [Weibel 2000a] WEIBEL, Stuart I.: **DC and Search Engines**. 16. Oktober 2000. – DC-GENERAL [dc-general@mailbase.ac.uk]
- [Weibel 2000b] WEIBEL, Stuart I. **The Dublin Core Metadata Initiative 8th Plenary Workshop**. 2000. – URL <http://purl.org/DC/workshops/dc8conference/pp/dc-8-plenary.ppt>. – Zugriffsdatum: 13.11.2000
- [Weibel und Lagoze 1997] WEIBEL, Stuart I. ; LAGOZE, Carl: **An element set to support resource discovery The state of the Dublin Core: January 1997**. In: *International Journal on Digital Libraries* 1 (1997), S. 176–186
- [Weibel und Miller 1997] WEIBEL, Stuart I. ; MILLER, Eric: **Image Description on the Internet - A Summary of the CNI/OCLC Image Metadata Workshop (September 24 - 25, 1996 Dublin, Ohio)**. In: *D-Lib Magazine* 3 (1997), Nr. 1. – URL <http://www.dlib.org/dlib/january97/oclc/01weibel.html>. – Zugriffsdatum: 28.10.2000
- [van der Werf-Davelaar 1999] WERF-DAVELAAR, Titia van d.: **Identification and resolution systems for networked access to information**. In: *Online and CD-ROM Review* 23 (1999), Dezember, Nr. 6. – URL <http://www.konbib.nl/persons/titia/publ/lij-v2.htm>. – Zugriffsdatum: 17.09.2000
- [Wood 1998] WOOD, Dee: **Online Peer Review?** In: *Learned Publishing* (1998), Nr. 11, S. 193–198