

HUMBOLDT-UNIVERSITÄT ZU
BERLIN

INSTITUT FÜR BIBLIOTHEKSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKSWISSENSCHAFT

HEFT 136

**DUBLIN CORE, SGML
UND DIE ZUKUNFT DER ERSCHLIESSUNG
AM BEISPIEL EINER STUDIE ZUR OPTIMIERUNG DES
DOKUMENTENMANAGEMENTS EINER GROSSEN
NICHTREGIERUNGSORGANISATION**

VON
GUDULA HOLZHEID

**DUBLIN CORE, SGML UND DIE ZUKUNFT DER
ERSCHLIESSUNG AM BEISPIEL EINER STUDIE ZUR
OPTIMIERUNG DES DOKUMENTENMANAGEMENTS
EINER GROSSEN NICHTREGIERUNGSORGANISATION**

VON
GUDULA HOLZHEID

Berliner Handreichungen
zur Bibliothekswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 136

Holzheid, Gudula:

Dublin Core, SGML und die Zukunft der Erschließung am Beispiel einer Studie zur Optimierung des Dokumentenmanagements einer großen Nichtregierungsorganisation / von Gudula Holzheid. - Berlin: Institut für Bibliothekswissenschaft der Humboldt-Universität zu Berlin, 2005. - 79 S. - (Berliner Handreichungen zur Bibliothekswissenschaft ; 136)

ISSN 1438-7662

Abstract:

Immer mehr Informationsobjekte werden in digitaler Form publiziert. In der vorliegenden Arbeit werden die Auswirkungen der Digitalisierung auf bibliothekarische Erschließungsmethoden dargestellt: einerseits die Funktion der Auszeichnungssprachen SGML, HTML und XML als Erschließungsinstrumente, andererseits der von bibliothekarischer Seite entwickelte Metadatenstandard Dublin Core. Am Praxisbeispiel "Dokumentenmanagement einer Nichtregierungsorganisation" wird untersucht, ob die Erschließung verbessert werden könnte durch z. B. Optimierung der Volltextindizierung, standardisierte Metadaten oder Integration von Metadaten und Suchmaschinentechologie. Mithilfe einer Benutzerbefragung werden diese Ansätze an der Praxis gemessen und konkrete Empfehlungen entwickelt.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang Master of Arts (Library and Information Science) an der Humboldt-Universität zu Berlin.

Inhalt	Seite
1. Einleitung	3
2. Auswirkungen der Digitalisierung auf bibliothekarisch-dokumentarische Erschließung	4
2.1. Wandel in der Funktion der Bibliothekskataloge, Regelwerksdebatte	4
2.2. Zunahme digitaler Erschließungsobjekte	6
2.3. Internet als Publikationsmedium	9
2.4. Erschließung von Internetressourcen	11
2.4.1. Auszeichnungssprachen SGML, HTML, XML: Definition und Abgrenzung.....	11
2.4.2. Funktion der Auszeichnungssprachen als Erschließungsinstrumente: Metadatenkonzept.....	13
2.4.3. „Semantic Web“	15
2.5. Dublin-Core-Metadatenstandard.....	17
2.5.1. Beschreibung, Definition	17
2.5.2. Anwendungsbeispiele	20
2.5.3. Einschätzung der zukünftigen Entwicklung.....	23
3. Praxisbeispiel: Dokumentenmanagement bei Greenpeace Deutschland.....	25
3.1. Hintergrund: Wissensmanagement	25
3.2. Dokumentenmanagement: Ausgangssituation, Problembeschreibung	26
3.3. Überschneidungsbereiche mit allgemeiner Erschließungsdiskussion...	29
3.3.1. Optimierung der Volltextindizierung	32
3.3.2. Dokumentenerschließung durch standardisierte Metadaten.....	35
3.3.3. Integration Metadaten/Suchmaschinentechnologie	39
3.4. Überprüfung der Ergebnisse durch Bedarfsanalyse: Benutzerbefragung	41
3.4.1. Inhalte der Befragung	41
3.4.2. Zielgruppe / NutzerInnen	42
3.4.3. Ergebnisse der Befragung	44
3.5. Empfehlungen / Lösungsvorschläge.....	47
3.5.1. Minimallösung.....	49
3.5.2. Maximallösung.....	50
4. Schlussbemerkung.....	51
5. Literaturverzeichnis.....	53

1. Einleitung

Bibliothekarische Erschließung – ein Zukunftsthema? Könnten nicht all die Veröffentlichungen und Kongresse rund um das Thema Metadaten nur Selbstvergewisserungsmaßnahmen von BibliothekarInnen sein, die angesichts der rasanten Entwicklungen im Internet die Bedeutung ihres Berufsstands als Wegweiser durchs Weltwissen schwinden sehen? Wird das ganze komplizierte „Gedöns“ nicht sowieso bald überflüssig, wenn immer leistungsfähigere Suchmaschinen uns in naher Zukunft genau die Fundstellen im weltweiten, mit Semantik aufgeladenen Internet zurückmelden, die wir für unser Thema suchen? Google ist längst schon Spitzenreiter bei den Recherchemethoden von Studierenden¹, diese „Abstimmung mit der Maus“ sollte doch hinsichtlich der Zukunftsfähigkeit bibliothekarischer Erschließung zu denken geben.

Oder nehmen wir das praktische Anwendungsbeispiel einer alles andere als lehrbuchmäßig erschlossenen Sammlung digitaler Dokumente bei einer großen Nichtregierungsorganisation, bei Greenpeace Deutschland. Nur sehr begrenzt wurde dokumentarische Sachkunde eingesetzt, um ein Erschließungssystem fürs elektronische Dokumentenmanagement zu ersinnen. Zusammen mit den bekannt eigensinnigen und auf selbständige, wenig hierarchisch strukturierter Arbeitsweise bestehenden MitarbeiterInnen ergab sich daraus mit den Jahren eine disparate, in Teilen durchaus chaotisch zu nennende Ansammlung digitaler Dokumente. Über 30 Datenbanken, über 100.000 Dokumente, kein konsistentes Erschließungssystem – das kann doch eigentlich nicht gutgehen, sagt jede/r dokumentarisch Sachkundige. Aber es funktioniert doch irgendwie, möglicherweise dank dem in die Datenbanken eingebauten leistungsfähigen Volltextindex. Irgendwie finden die Greenpeace-MitarbeiterInnen die Dokumente, die sie suchen. Ob es genau die sind, die sie „eigentlich“ suchen, und ob nicht schon viele Schätze des Greenpeace-Gedächtnisses auf diese Weise verschüttet wurden, ist eine andere Frage.

In diesem Spannungsfeld liegt die Motivation dieser Arbeit: Einerseits ist sicherlich jenseits aller Polemik im BID-Bereich die Aussage konsensfähig, dass durch die aktuellen Entwicklungen rund um digitale Publikationen die klassischen bibliothekarischen Erschließungsmethoden, die sich Jahrtausende lang auf analoge Medien bezogen, zumindest ergänzungsbedürftig sind. Andererseits besteht bei der Erschließung digitaler Dokumente – seien sie im Internet publiziert oder im Intranet einer Nichtregierungsorganisation in internen Datenbanken eines Dokumentenmanagementsystems gespeichert – sicherlich über die heutzutage selbstverständliche Volltextindizierung und Suchmaschinensuche hinaus noch viel Verbesserungsbedarf, um den BenutzerInnen Orientierung bei der Suche nach relevanten Dokumenten zu bieten. Anders formuliert: Möglicherweise könnten die klassischen Funktionen der bibliothekarisch-dokumentarischen Erschließung unter den Bedingungen der Digitalisierung wiederzuentdecken sein. Das Praxisbeispiel „Dokumentenmanagement einer Nichtregierungsorganisation“ wurde gewählt, weil es inhaltlich als Teil eines Wissensmanagementsystems zu

¹ In der so genannten SteFi-Studie („Studieren mit elektronischen Fachinformationen“) des Bundesministeriums für Bildung und Forschung wurde 2001 festgestellt, dass fast zwei Drittel der Studierenden elektronische Fachinformationen mithilfe kommerzieller Internetsuchmaschinen suchen, siehe <http://www.stefi.de/download/kurzfas.pdf> (Stand vom 12. 05. 2004)

bewerten ist. Wissensmanagement in Unternehmen und Organisationen benutzt inzwischen Methoden und Verfahren, die bislang auf Literaturdokumentation in Fachdatenbanken oder auf Informationsgewinnung aus dem Internet beschränkt waren (Nohr 2003:12). Deshalb erscheint es naheliegend, dass in der gegenwärtigen Fachdiskussion und Praxis bibliothekarisch-dokumentarischer Erschließung Probleme thematisiert und möglicherweise sogar gelöst werden, die auch im Praxisbeispiel vorliegen. Zunächst ist also zu fragen, welche Tendenzen die derzeitige Situation der Erschließung im BID-Bereich kennzeichnen, welche Themen zur Zeit die Fachdiskussion beherrschen, dann werde ich die Problemstellung im Praxisbeispiel Dokumentenmanagement bei Greenpeace Deutschland genauer beschreiben und versuchen, Überschneidungsbereiche zu isolieren, die dann im Detail diskutiert werden. Letztlich wird der so entwickelte theoretische Lösungsansatz durch eine Benutzerbefragung an der Praxis gemessen und mündet in konkrete Empfehlungen.

2. Auswirkungen der Digitalisierung auf bibliothekarisch-dokumentarische Erschließung

Die gegenwärtige Situation in der bibliothekarisch-dokumentarischen Erschließungspraxis und der darüber geführten Fachdiskussion ist vor allem durch die vielfältigen Auswirkungen der Digitalisierung gekennzeichnet. Einerseits hielt die elektronische Datenverarbeitung in den Bibliotheken Einzug, was zu einem Wandel in der Funktion der Kataloge führte, andererseits sind zunehmend nicht nur analoge, sondern auch digitale Informationsobjekte zu erschließen.

2.1. Wandel in der Funktion der Bibliothekskataloge, Regelwerksdebatte

Seit den 1960er Jahren wurden in Deutschland die klassischen Bibliothekskataloge durch Systeme elektronischer Datenverarbeitung ersetzt (Hoffmann/Diedrichs 2004:418), und seitdem ist eine Konvergenz von Formal- und Sacherschließung zu beobachten.

Traditionell wird im bibliothekarischen Bereich zwischen Formal- und Sacherschließung unterschieden. Formalerschließung hat zum Ziel, eine Suche nach bekannten Informationsobjekten zu ermöglichen, z.B. eine Titelseuche. Im alphabetischen Bibliothekskatalog werden die Ergebnisse der Formalerschließung niedergelegt, der Katalog ist also einerseits Bestandsnachweis, andererseits Suchhilfsmittel. Sacherschließung dagegen soll eine themenzentrierte Recherche ermöglichen, also alle Dokumente zu einem gegebenen Thema oder Problem finden helfen, auch wenn formale Eigenschaften des Dokuments wie Titel, Autor etc. unbekannt sind. Dazu mussten vor der Einführung digitaler Kataloge zusätzliche Register mit Schlagwörtern oder anderen inhaltlichen Erschließungsdaten angelegt werden, deren Einträge auf formal erschlossene Dokumente verwiesen.

Seit sich digitale Kataloge in Bibliotheken etabliert haben, kann auch in Daten, denen lediglich eine Formalerschließung zugrunde liegt, nach dem Vorkommen

bestimmter Stichwörter gesucht werden. Naheliegenderweise bieten sich hierfür vor allem die Titel der Informationsobjekte an. In der Praxis ergänzen sich bei Themenrecherchen Stichwortsuche in den Titelfeldern und Schlagwortsuche gegenseitig, die Abgrenzung ist für NutzerInnen nicht mehr nachvollziehbar (Haller/Fabian 2004:226). In der heutigen Bibliothekslandschaft sind Formal- und Sacherschließung durch die Existenz digitaler Kataloge also faktisch zusammengewachsen, Bibliothekskataloge haben jetzt nicht mehr nur die Funktion von Bestandsnachweisen, sondern auch von Suchwerkzeugen für thematische Stichwortsuchen.

Als weitere Konsequenz der Digitalisierung der Kataloge wurde der Datenaustausch zwischen den verschiedenen Akteuren im BID-Bereich ermöglicht. So entstanden in Deutschland seit den 70er Jahren Bibliotheksverbundsysteme, in denen Datenbanksysteme zur verteilten Erfassung und gemeinsamen Nutzung von Katalogdaten entwickelt wurden. In den USA entstand mit OCLC das wohl weltweit größte Verbundsystem (Hoffmann/Diedrichs 2004:444). Heute sind über 45.000 Bibliotheken in 84 Staaten durch OCLC zusammengeschlossen (Jordan 2004). Es liegt nahe, dass eine derart große, global agierende Institution eine eigene Dynamik in der internationalen Fachwelt entfaltet, dass ein Druck auf kleinere nationale Systeme entsteht, ihre Standards möglichst leicht austauschbar, also interoperabel mit denen von OCLC zu machen. Interoperabilität zwischen den bisher sehr unterschiedlichen Regelwerken AACR und RAK/RWSK ist nur um den Preis erhöhten Aufwands für Konvertierungen, komplexe Datenaustauschformate etc. machbar. Führt man sich vor Augen, dass der Aufwand für Softwareentwicklung im Bibliotheksbereich immens ist, da einerseits in den großen Staats- und Universitätsbibliotheken Massendaten zu bewältigen sind, andererseits die Software nur in geringer Stückzahl verkauft werden kann, so wird deutlich, dass mit erhöhtem Kostendruck auf die Bibliotheken die so genannte „Regelwerksdebatte“ nicht ausbleiben konnte. Darin wurde die Übernahme der angloamerikanischen Regelwerke in die Bibliothekspraxis im deutschsprachigen Raum diskutiert, also die Frage: AACR oder RAK/RWSK?

Die Debatte wurde nach dem „Nikolausbeschluss“ Ende 2001 in der bibliothekarischen Fachöffentlichkeit sehr engagiert, teilweise sogar mit einer für den Berufsstand sonst eher untypischen Leidenschaftlichkeit geführt². Am 6. Dezember 2001 hatte der Standardisierungsausschuss der Deutschen Bibliothek entschieden, dass die Regeln für die alphabetische Katalogisierung nicht mehr weiterentwickelt werden sollen – was implizit eine Entscheidung für die Übernahme der US-amerikanischen Regelwerke bedeutete. Dieser Beschluss wurde zwar im November 2002 teilweise zurückgenommen, jedenfalls insoweit entschärft, dass die deutschsprachigen Regelwerke weiterentwickelt werden sollen, bis eine endgültige Entscheidung gefällt wird. Diese endgültige Entscheidung wird zur Zeit durch vielfältige Projektaktivitäten der Deutschen Bibliothek vorbereitet, so wurde die Unternehmensberatung Kienbaum mit einer Wirtschaftlichkeitsuntersuchung beauftragt, deren Endbericht seit Ende 2003 vorliegt.³ Eine US-amerikanische Unternehmensberatung hat in einem Teilprojekt eine Studie angefertigt und kommt erwartungsgemäß zum Ergebnis, dass die

² siehe die von der Universitätsbibliothek Augsburg zusammengestellten Links:
http://www.bibliothek.uni-augsburg.de/kfe/inf_rwd.html (Stand vom 12. 05. 2004)

³ http://www.bibliothek.uni-augsburg.de/kfe/mat/endber_wirt.pdf (Stand vom 12. 05. 2004)

Übernahme internationaler Standards anzuraten sei⁴. Die Regelwerksdebatte in der deutschen bibliothekarischen Fachöffentlichkeit ist inzwischen zur Ruhe gekommen, die Übernahme der angloamerikanischen Datenaustauschformate für Katalogisate scheint nur noch eine Frage der Zeit zu sein, zunächst müssen allerdings noch viele offene Fragen nach einem Maßnahmenkatalog, nach Kosten und dem Zeithorizont geklärt und entschieden werden (Krause/Niggemann/Schwänzl 2003:21).

Um zukünftige Entwicklungstendenzen bei der bibliothekarisch-dokumentarischen Erschließung im deutschsprachigen Raum zu beurteilen, müssen also auch absehbare zukünftige Erschließungskonzepte im angloamerikanischen Bereich einbezogen werden. Dort scheint zur Zeit ein Entitätenkonzept favorisiert zu werden, wie es in den Functional Requirements for Bibliographic Records (FRBR)⁵ von der International Federation of Library Associations and Institutions (IFLA) konkretisiert wurde. Es handelt sich um eine hierarchische Struktur, in der Informationsobjekte in ihren Ausprägungen als „work / expression / manifestation / item“ unterschieden werden. Dieses Konzept wird in Australien und Skandinavien teilweise schon angewendet und scheint sich für die Systematisierung analoger Erschließungsobjekte als zukunftsweisender Ansatz zu erweisen. Ob es allerdings für digitale Informationsobjekte brauchbar ist, muss bezweifelt werden. Caplan findet instruktive Beispiele, warum das Entitätenkonzept sich für digitale Informationsobjekte von seinem Ansatz her nicht eignet:

„Electronic resources also challenge FRBR. A manifestation is defined as all copies in the same medium in the same physical form, but what constitutes identity of a medium is left unexplored. ... Is the same physical disk or tape a different medium if it is online or offline? Common sense tells us that there is a level of discrimination beyond which metadata creators are unlikely to know, and users unlikely to care, about differences in media.“

(Caplan 2003:10)

2.2. Zunahme digitaler Erschließungsobjekte

Nicht nur auf Seiten der Bibliothekstechnik, sondern auch vor allem in Bezug auf die zu erschließenden Informationsobjekte hat die Digitalisierung zu großen Veränderungen im BID-Bereich geführt. Immer häufiger sind die Objekte bibliothekarisch-dokumentarischer Erschließung nicht mehr die klassischen analogen Medien wie Monographien oder Zeitschriften, sondern liegen in digitaler Form vor. Das nimmt nicht wunder, führt man sich vor Augen, dass im Jahr 2002 über 90 Prozent der Flut neuer Information weltweit in elektronischer Form erzeugt und gespeichert wurde, wie die „How much information“-Studie⁶ ergab. Auch wenn ein großer Teil der dort untersuchten Information nicht als publizierte

⁴ „Changing to international standards could have many short-term and long-term benefits for German libraries, including simplified record transfer and sharing, increased efficiency in cataloging, and the opportunity to participate in ongoing research and development on new standards“, stellt die Unternehmensberatung Stillwater Group in ihrem Abschlussbericht für Die Deutsche Bibliothek am 8. 3. 2004 fest. http://www.bibliothek.uni-augsburg.de/kfe/mat/final_report.pdf (Stand vom 12. 05. 2004)

⁵ <http://www.ifla.org/VII/s13/frbr/frbr.pdf> (Stand vom 12. 05. 2004)

⁶ Lyman/Varian: How Much Information, 2003.

<http://www.sims.berkeley.edu/research/projects/how-much-info-2003> (Stand vom 12. 05. 2004)

Fachinformation anzusehen ist, ist die Tendenz weg von analogen und hin zu originär digitalen („born digital“) Informationsobjekten eindeutig.

In den 1960er Jahren begann diese Entwicklung mit der Entstehung elektronischer Fachinformation im dokumentarischen Bereich: Fachinformationsdienste wie z.B. „Current Contents“ oder „Chemical Abstracts“ wurden digitalisiert und standen dann als Datenbanken auf Großrechnern („hosts“) zu Recherchen mit spezifischen Abfragesprachen zur Verfügung. Inhaltlich handelte es sich dabei um sekundäre Informationsobjekte („abstracts“), denen noch die primären Publikationen als Zeitschriftenartikel in analoger Form zugrunde lagen, es waren also eigentlich Ergebnisse dokumentarischer Erschließungstätigkeit, die ihrerseits durch das Datenbanksystem mit seinen inhärenten Abfragesprachen erschlossen waren. Später waren bestimmte Datenbankinhalte dann auch per Volltextindex erschlossen, so dass Recherchen beispielsweise auf die Volltexte der Abstracts möglich waren.

Mit immer leistungsfähigeren Rechnern, Speicher- und Übertragungsmedien werden nun auch zunehmend primär digitale Informationsobjekte von Institutionen im BID-Bereich erschlossen. Viele wissenschaftliche Zeitschriften werden parallel zur Printausgabe oder inzwischen auch ausschließlich in digitaler Form publiziert, Dissertationen werden auf Universitätsservern veröffentlicht und in Bibliothekskatalogen verzeichnet, Inhalte von Lehrveranstaltungen werden als „digitaler Semesterapparat“ von Universitätsbibliotheken angeboten etc. Damit entstehen vielfältige neue Herausforderungen für die erschließende Institution.

Als erstes fällt auf: Die Granularitätsebene ändert sich, es werden nicht mehr wie bisher im Bibliotheksbereich meist üblich nur Monographien erschlossen, sondern auch Dokumente, die klassischerweise als unselbständige Literatur galten. Interessant ist in diesem Zusammenhang der historische Vergleich zum Aufkommen der Zeitschriften im 17. Jahrhundert. Einerseits wird heute in der Fachliteratur die Meinung vertreten, die gegenwärtige, durch die Digitalisierung bewirkte Umbruchsituation im BID-Bereich sei ähnlich tiefgreifend wie diejenige im 17. Jahrhundert. Andererseits wird auch daran erinnert, dass ein damals entstandenes Problem der Erschließung – nämlich die Erschließung unselbständig erschienener Zeitschriftenaufsätze – bis heute in Bibliotheken nicht zufriedenstellend gelöst ist. Als „keine tröstende Ausgangsposition“ erschien dies Hermann Leskien schon 1997 (Leskien 1997:1). Umso mehr muss genau überprüft werden, ob die hergebrachten Erschließungsmethoden für die neu entstandenen digitalen Informationsobjekte tauglich sind, oder ob alternative Konzepte entwickelt und angewendet werden sollten.

Die Volltextindizierung des gesamten Inhalts der Informationsobjekte ist durch die Fortschritte in der Informationstechnik inzwischen nicht nur möglich, sie ist in vielen Anwendungsgebieten sogar selbstverständlich geworden. Daraus entstand eine völlig neue Situation im Erschließungsbereich: Die „Reduktion auf Wesentliches“, die intellektuelle Erschließung, wurde teurer als das Anbieten aller Textwörter als Suchbegriffe (Knorz 1997:130). Es stellte sich nun die Frage, ob der Aufwand intellektueller Indizierung noch gerechtfertigt ist, oder ob nicht vielmehr automatisierte Indexierungsverfahren die kostengünstigere und sogar effektivere Alternative sein könnten. Mitte der 90er Jahre wurde diese Diskussion sehr engagiert in der Fachöffentlichkeit geführt (Dörr/Enderle/Hauße 2004:401),

ihre Inhalte sind inzwischen aufgegangen in der Diskussion über Metadaten und Suchmaschinentechnologie, siehe dazu unten Abschnitt 2.5.3.

Digitale Dokumente sind leicht veränderbar, jedenfalls leichter als in gedruckter Form vorliegende analoge Medien. Daraus entstehen Probleme der Authentizität und Versionskontrolle: Wie kann sichergestellt werden, dass das Dokument tatsächlich von AutorIn XY stammt und nicht verändert wurde (Gladney/Bennett 2003)? Als Lösungsansätze kommen Systeme digitaler Signaturen und Authentisierungs-codes in Betracht, die im Bereich des öffentlichen Internets entwickelt wurden und auch in digitalen Bibliothekssystemen angewendet werden, wenn die Bibliothek selbst als publizierende Institution tätig wird, was ebenfalls durch die Digitalisierung befördert wurde. Als Beispiel könnte hier der Dokumentenserver der Deutschen Bibliothek dienen, auf dem seit 1998 Dissertationen und Habilitationen gesammelt werden,⁷ deren Authentizität und Integrität mit aufwändigen Verfahren gesichert werden.

Digitale Dokumente benötigen Hard- und Software, um genutzt werden zu können. Daraus ergibt sich eine neue Funktion von Erschließung, nämlich zu Zwecken der Langzeitarchivierung. Der rasche Generationswechsel bei Hard- und Software bedingt, dass heute noch weitverbreitete Formate morgen schon veraltet sind. Kritisch wird es, wenn keine Hard- oder Software mehr existiert, mit der veraltete Dokumente genutzt werden können. Im Rahmen der Erschließung müssen für digitale Dokumente also Informationen über die technischen Nutzungsvoraussetzungen erhoben und gespeichert werden, weil sonst jeder Ansatz der Langzeitarchivierung zum Scheitern verurteilt wäre. Archivierungsstrategien werden zur Zeit vielfältig erdacht und erforscht, sowohl in technischer als auch in organisatorischer Hinsicht (Dobratz/Liegmann/Tappenbeck 2001:326 ff.), die Deutsche Bibliothek hat im Projekt „nestor – Kompetenznetzwerk Langzeitarchivierung“⁸ die Federführung für Deutschland übernommen, auf internationaler Ebene ist in erster Linie das Projekt Open Archives Initiative⁹ zu nennen. Trotz oder gerade wegen der vielfältigen Aktivitäten ist zur Zeit noch kein endgültiger Stand der Diskussion absehbar, es würde den Rahmen dieser Arbeit sprengen, die verschiedenen Ansätze darzustellen und fundiert zu kommentieren. Festzustellen bleibt, dass zur Zeit noch kein einsatzfähiges Verfahren oder auch nur eine allgemein anerkannte Vorgehensweise existiert, um die Probleme der Langzeitarchivierung digitaler Dokumente zu lösen.

Deshalb wird der Erschließungsbegriff im Rahmen dieser Arbeit auf Erschließung für zeitnahen Zugriff eingegrenzt, mögliche Erschließungsfunktionen im Rahmen von Langzeitarchivierung werden ausgeklammert, allerdings nur in dem Bereich, der sich mit den technischen Problemen der Langzeitarchivierung digitaler Dokumente beschäftigt. Im inhaltlichen Kernbereich der Erschließung, wenn es zum Beispiel um die Dynamik des zu erschließenden Themas geht, ist auch im Rahmen dieser Arbeit im Blick zu behalten, dass Erschließung immer zukunftssicher sein muss.

⁷ <http://www.ddb.de/professionell/dissonline.htm> (Stand vom 12. 05. 2004)

⁸ <http://www.langzeitarchivierung.de/index.php> (Stand vom 12. 05. 2004)

⁹ <http://www.openarchives.org/> (Stand vom 12. 05. 2004)

Letztlich bleibt noch ein weiterer Effekt zu nennen, den die Zunahme digitaler Informationsobjekte auf die klassische bibliothekarisch-dokumentarische Erschließung hatte: Die Rollen der Akteure sind nicht mehr so klar voneinander zu unterscheiden wie bisher. Wer erschließt für wen? Die traditionell scharf getrennten Rollen Autor/Dokumentar/Nutzer scheinen zu konvergieren. Ein gutes Beispiel bietet das UrMEL-Projekt¹⁰: Hier wurde eine digitale Multimedia-Universitätsbibliothek realisiert, die Erschließung der Dokumente ist in einem „Geschäftsgang Digitale Dokumente“ in einer Work-flow-Anwendung organisiert (Kuhles/Wefers 2001:224). Bestimmte Dokumenttypen, z.B. Dissertationen, werden durch die AutorInnen selbst erschlossen, sie füllen ein elektronisches Erfassungsformular aus. In einem zweiten Schritt prüfen und ergänzen BibliotheksmitarbeiterInnen die Erschließungsdaten. Die Bibliothek hat hier also die Funktion einer Qualitätssicherung übernommen.

Wie insgesamt im BID-Bereich ist also auch im Bereich der Erschließung eine Umbruchsituation zu beobachten, die in der Digitalisierung ihren Ursprung hat. Einige klassische Denkmodelle und Handlungsmuster, die für analoge Medien sinnvoll und richtig waren, scheinen unter den gegenwärtigen Bedingungen nicht mehr der Realität zu entsprechen und müssen analysiert und bei Bedarf verändert werden. Zusammenfassend können folgende Themenbereiche aufgezählt werden, die im Zusammenhang mit der Erschließung digitaler Informationsobjekte in der Fachdiskussion aufgetaucht sind:

- Volltextindizierung
- Authentizität und Versionskontrolle
- Langzeitarchivierung (im Rahmen dieser Arbeit nicht ausgeführt)
- Rollenkonvergenz der Akteure

2.3. Internet als Publikationsmedium

Internet-Quellen gewinnen als Basis wissenschaftlichen Arbeitens zunehmend an Bedeutung (Dörr/Enderle/Hauße 2004:399). Seit den frühen 90er Jahren wird das Internet immer mehr zur Publikation von Dokumenten aus Wissenschaft und Forschung genutzt. Auf so genannten Preprint- oder Eprint-Servern werden wissenschaftliche Arbeiten publiziert, als Beispiel sei der Physik-Server arXiv¹¹ herausgegriffen, der vom Los Alamos National Laboratory¹² betrieben wird. Seit 1991 werden dort Forschungsdokumente („research papers“) aus Physik und verwandten Fachgebieten veröffentlicht,¹³ jedes Jahr kommen etwa 40.000 Dokumente hinzu. Daran wird deutlich, dass im Internet keine klar definierte Grenze zwischen wissenschaftlicher Kommunikation und wissenschaftlicher Publikation existiert, genausowenig wie zwischen allgemeiner oder kommerzieller Information auf der einen und Fachinformation auf deren anderen Seite (Dörr/Enderle/Hauße 2004:400).

¹⁰ University Multimedia Electronic Library an der Thüringer Universitäts- und Landesbibliothek Jena

¹¹ <http://xxx.lanl.gov/> (Stand vom 12. 05. 2004)

¹² eine der größten US-amerikanischen Forschungsinstitutionen, siehe <http://www.lanl.gov/organization/> (Stand vom 12. 05. 2004)

¹³ „Started in August 1991, arXiv.org (formerly xxx.lanl.gov) is a fully automated electronic archive and distribution server for research papers. Covered areas include physics and related disciplines, mathematics, nonlinear sciences, computational linguistics, and neuroscience.“ <http://lanl.arxiv.org/help/general> (Stand vom 12. 05. 2004)

Kommerziell betriebene Suchmaschinen bieten die Volltextindizierungssysteme, mit denen „das Internet“ zur Zeit am ehesten als erschlossen bezeichnet werden kann. Die Datenbankinhalte der Suchmaschinen werden meist durch so genannte Robots automatisch generiert, dazu werden Webserver beispielsweise durch Linkverfolgung rekursiv abgesucht und Informationen zu den gefundenen Dokumenten abgespeichert. Diese Informationen sind zunächst unstrukturiert, was zu einer geringen Qualität im Retrievalprozess führt: zu hohe Trefferzahl, zu geringe precision-Rate und Schwierigkeiten bei der Selektion und Bewertung der Ergebnisse¹⁴. Die ungeheure Heterogenität der im Internet vorhandenen Dokumente in Hinblick auf Inhalt, Format, Qualität, Spezifität, Aktualität, Nutzungsziele und Zielgruppen (Ferber 2002:931) ist also das erste Problem, das sich NutzerInnen stellt, die mithilfe kommerzieller Suchmaschinen versuchen, gezielt nach Fachinformationen zu suchen. Aber auch wenn man die sprichwörtliche Nadel im Heuhaufen gefunden hat, taucht ein weiteres, nicht minder schwerwiegendes Problem auf: Man kann nicht sicher sein, dass die mühsam gefundene Informationsquelle morgen immer noch unter der angegebenen Adresse zu finden ist.

Dieses Problem der Persistenz von Internetadressen resultiert aus der dezentralen Struktur des Internets und hat seinen Ursprung in der Organisation der Internetadressierung. Für wissenschaftliche Fachinformation ist Persistenz aber unabdingbar, weil wissenschaftliches Arbeiten Nachprüfbarkeit der Quellen voraussetzt.

Zur Zeit existieren drei Lösungsansätze, mit denen unterschiedliche Organisationen versuchen, Systeme dauerhafter Internetadressierung aufzubauen:¹⁵

- Digital Object Identifier (DOI): ist zur Zeit das einzige System, das in praktischen Anwendungen produktiv verwendet wird. Es geht auf eine Initiative des nordamerikanischen Verlegerverbands (Association of American Publishers) zurück, erlaubt die Zuordnung von Nutzungs- und Lizenzrechten zu eindeutig identifizierten Dokumenten und kann deshalb als Basis für Abrechnungssysteme dienen.
- Uniform Resource Name (URN): wurde von der Internet Engineering Task Force¹⁶ entwickelt und soll die orts- und rechnertransparente Adressierung ermöglichen. So genannte Agenturen sollen als Verwaltungsinstanzen eindeutige Dokumenten- oder Ressource-Bezeichner verwalten, die an Webseiten-Betreiber vergeben werden. Auf Initiative der Conference of Directors of National Libraries wird diskutiert, ob diese Agenturen bei den jeweiligen Nationalbibliotheken angesiedelt werden könnten.
- Persistent Uniform Resource Locator (PURL): ähnelt dem URN-Konzept und wurde von OCLC entwickelt.

¹⁴ zur Evaluierung von Suchsystemen vgl. zusammenfassend Nohr 2003:119 ff. und Ferber 2002:926 ff.

¹⁵ vgl. die Zusammenstellung bei Dörr/Enderle/Hauße 2004:412

¹⁶ IETF, ein Zusammenschluss von Einzelpersonen aus Forschung und Technik und vor allem großen Firmen wie Cisco, IBM etc., der die Entwicklung neuer Standards für die im Internet benutzten Technologien vorantreiben will, siehe <http://www.ietf.org/overview.html> (Stand vom 12. 05. 2004)

Welcher dieser Lösungsansätze sich durchsetzen wird, ist zur Zeit noch nicht absehbar. Es würde den Rahmen dieser Arbeit sprengen, Vor- und Nachteile der Konzepte abzuwägen, und einzuschätzen, welcher der beteiligten Institutionen es gelingen könnte, die notwendige kritische Masse von Anwendungen ihres Konzepts zu erreichen.

2.4. Erschließung von Internetressourcen

Auch wenn im Problembereich der Persistenz also noch kein endgültiger Lösungsansatz absehbar ist, so gibt es doch vielfältige Aktivitäten in der Fachöffentlichkeit, um den oben geschilderten Problemen der Erschließung von Dokumenten im anarchischen, unstrukturierten Internet zu begegnen. Es sind Initiativen aus zwei unterschiedlichen fachlichen und institutionellen Richtungen entstanden: einerseits wird von Seiten des World-Wide-Web-Konsortiums W3C¹⁷ versucht, die im Internet vorhandene Technologie der Auszeichnungssprachen weiterzuentwickeln und als Erschließungsinstrument zu nutzen; auf der anderen Seite werden im bibliothekarischen Bereich Metadatenstandards erarbeitet, mit deren Hilfe die Funktionen klassisch bibliothekarisch-dokumentarischer Erschließung auch für im Internet publizierte Dokumente annähernd erreicht werden sollen, also NutzerInnen Orientierung auf der Suche nach relevanter Fachinformation geboten werden könnte.

2.4.1. Auszeichnungssprachen

SGML, HTML, XML: Definition und Abgrenzung

Die überwältigende Mehrheit aller zur Zeit im Internet vorhandenen Informationsobjekte sind Textdokumente im HTML-Format, sie wurden also unter Verwendung der Auszeichnungssprache Hypertext Markup Language erstellt (Schiffer/Tempel 2002:1016). Welche Implikationen könnte dies für ihre Erschließung haben?

Auszeichnungssprachen sind Ende der 1960er Jahre ursprünglich entstanden, um existierende Standards für die typographische Auszeichnung analoger Dokumente auch für die Gestaltung digitaler Dokumente nutzen zu können. 1969 entwickelte der bei IBM angestellte Jurist Charles Goldfarb zusammen mit Edward Mosher und Raymond Lorie die erste generische Auszeichnungssprache Generalized Markup Language (GML), deren Abkürzung nicht zufällig aus den Initialen der drei IBM-Mitarbeiter besteht. Im Rahmen eines Forschungsprojekts über integrierte Informationssysteme für die juristische Praxis sollte GML dazu dienen, Dokumente zwischen den Subsystemen für Textverarbeitung, Textformatierung und Information Retrieval austauschbar zu machen. Im Gegensatz zu den damals schon bestehenden typographischen Auszeichnungssprachen für Zwecke der Druckindustrie basierte GML nicht nur auf einem einfachen System von „tags“ (generischen Befehlen, um unabhängig von der konkreten Umsetzung in Schriftarten, Schriftgrößen etc. beispielsweise eine Überschrift zu kennzeichnen), sondern verwendete erstmals das Konzept eines formal definierten Dokumenttyps

¹⁷ Standardisierungsgremium für die technischen Grundlagen des Internets, siehe <http://www.w3.org/Consortium/> (Stand vom 12. 05. 2004); zur Abgrenzung von Standardisierung und Normung vgl. Kamke/Zimmermann 2003:345

mit einer explizit dazugehörigen Elementstruktur (SGML Users' Group 1990, Goldfarb 1996).

Goldfarb entwickelte GML zu SGML weiter, allerdings nicht mehr unter dem Dach von IBM, sondern im Rahmen eines eigens zu diesem Zweck gegründeten Ausschusses des American National Standards Institute, das 1980 den ersten funktionsfähigen Entwurf des SGML-Standards publizierte. Seit 1986 ist SGML in weiterentwickelter Form auch ein internationaler Standard unter der Kennzeichnung ISO 8879:1986¹⁸, dem die DIN-Norm DIN EN 28879¹⁹ entspricht.

Die Norm spezifiziert eine Sprache zur Repräsentation von Dokumenten und basiert auf einem Bearbeitungsmodell, bei dem Dokumente als Struktur verschiedener Typen von Elementen betrachtet werden (Wende 2002:1122). Sie ist insofern als Metasprache zu bezeichnen, als sie Dokumentenbeschreibungssprachen beschreibt (Caplan 2003:17).²⁰

Folgende Systematisierung macht das Prinzip anschaulich:

SGML beschreibt

- Dokumenttyp-Definitionen (SGML-DTDs), diese enthalten jeweils zulässige
 - Dokumentstrukturen und
 - Auszeichnungsbefehle („markups“ oder „tags“)
- Formatierungsstile mithilfe der Sprache DSSSL (Mühlhäuser 2002:874).

Kennzeichnend für das Konzept von SGML ist damit, dass Inhalt, Struktur und Gestaltung bzw. Formatierung von Dokumenten unabhängig voneinander definiert, gespeichert und weiterbearbeitet werden können. Dieses sehr einleuchtende Konzept hat allerdings den Preis einer hohen Komplexität, worin wahrscheinlich auch der Grund zu sehen ist, dass nicht SGML, sondern die davon abgeleitete Dokumentenbeschreibungssprache HTML (Hyper Text Markup Language) sich als Format für die übergroße Mehrheit der im World Wide Web vorhandenen Dokumente durchgesetzt hat.

HTML wurde ebenfalls durch eine ISO-Norm²¹ standardisiert. Es ist eine Ausprägung von SGML, eine Dokumenttyp-Definition (DTD), die allerdings das für SGML fundamentale Prinzip der Trennung von Struktur und Formatierung nicht befolgt. Dies wird in Literatur und Praxis allgemein als große Schwachstelle beklagt (z.B. Mühlhäuser 2002:874), hat aber den Komplexitätsgrad in der Anwendung der Auszeichnungssprache entscheidend verringert. Ein weiterer Nachteil von HTML-Anwendungen liegt darin, dass Verknüpfungen auf andere Dokumente oder auf Teile desselben Dokuments in ein HTML-Dokument eingebettet sind, statt in einer getrennten Schicht angegeben zu sein, was zu einer Inkonsistenz der Verknüpfungen (einerseits innerhalb desselben HTML-Dokuments, andererseits auf Ziele außerhalb des Dokuments) führt und damit das

¹⁸ ISO 8879:1986: Information processing - Text and office systems - Standard Generalized Markup Language (SGML). 1986.

¹⁹ DIN EN 28879 Informationsverarbeitung; Textverarbeitung und -kommunikation; genormte verallgemeinerte Auszeichnungssprache (SGML)

²⁰ „SGML is formally a metalanguage, or a language for describing other languages.“

²¹ ISO/IEC 15445: Informationstechnik; Dokumentbeschreibungs- und -verarbeitungsprachen; Hyper Text Markup Language (HTML)

oben beschrieben Persistenzproblem verschärft, weil es Rückwärts-Auflösungen von Verknüpfungen nicht in allen Fällen zulässt.

Als Reaktion auf die bei HTML zutage getretenen Unzulänglichkeiten auf der einen Seite, und angesichts der enormen Komplexität und damit geringen Verbreitung von SGML andererseits, entwickelte sich quasi als Nachfolgesprache von SGML seit Mitte der 90er Jahre die Extensible Markup Language (XML). Sie ist wie SGML eine Metasprache, aber von der Syntax her einfacher aufgebaut. XML besteht aus einer Teilmenge des SGML-Standards, so dass es als abgespeckte Version von SGML gelten kann (Mühlhäuser 2002:875). Es wird auch als Teilmenge von SGML bezeichnet, die striktere Regeln, weniger Funktionalität und weniger Flexibilität bietet, aber dafür sowohl von Menschen als auch von Maschinen einfacher zu handhaben ist (Caplan 2003:19)²².

Die Entwicklung und Verbreitung der neuen Auszeichnungssprache wird sowohl vom World-Wide-Web-Konsortium W3C als auch von den Normungsinstitutionen²³ betrieben (Wende 2002:1123). Die Initiative scheint erfolgreich zu sein, zwar ist heute noch die übergroße Mehrheit der im Internet vorhandenen digitalen Objekte mit HTML-Editoren erzeugt worden, aber XML erfreut sich zunehmender Beliebtheit, wie Umstätter schon im Jahr 2000 feststellte (Umstätter 2000:249); immer mehr Internetressourcen benutzen heute diese Auszeichnungssprache. „XML is ... becoming the language of the Web“, schreibt Caplan (2003:20). Bemerkenswert ist außerdem, dass Microsoft in der neuesten Version seiner Textverarbeitung die Dokumente nicht mehr im proprietären Microsoft-Word-Format, sondern in XML erzeugt. Auch dies lässt hoffen, dass XML tatsächlich in den nächsten Jahren ein universelles, frei verfügbares Dokumentenstandard-Format werden wird.

Ein Vorteil von XML ist, dass es erweiterbar („extensible“) ist. Um trotzdem Interoperabilität zu gewährleisten, werden nur bestimmte Dokumenttypen von W3-Konsortium anerkannt. Noch höher ist aber der Vorteil einzuschätzen, dass mithilfe von XML auch so genannte semantische „tags“ realisiert werden können.

2.4.2. Funktion der Auszeichnungssprachen als Erschließungsinstrumente: Metadatenkonzept

Prinzipiell liegt die wichtigste Funktion der Auszeichnungssprachen als Erschließungsinstrumente in der damit möglichen Trennung von Form, Struktur und Inhalt. Mit ihrer Hilfe können bisher unstrukturierte digitale Informationsobjekte erschlossen werden in dem Sinn, dass ihr Inhalt unabhängig von seiner Form beschrieben werden kann, und zwar mithilfe einer inhärenten Strukturierung, in die Metadaten integriert werden können.

Metadaten sind dem Wortsinn nach Daten über Daten oder konkreter: „maschinenlesbare Daten über Internetressourcen und andere Dinge“²⁴, so die

²² „XML can be thought of as a subset of SGML design with tighter rules, fewer features, and fewer options, in order to make it easier to process.“

²³ ISO/IEC JTC 1/SC 34: Extensible Markup Language (XML)

²⁴ „Metadata is machine understandable information about web resources or other things“, Berners-Lee 1997

Definition von Tim Berners-Lee, einem der Pioniere des World Wide Web und heutigen Direktor des W3-Konsortiums. Metadaten sind also strukturierte Daten, mit deren Hilfe Informationsobjekte beschrieben werden, indem sie beispielsweise Angaben über Titel, Autor oder Zeitpunkt der Veröffentlichung enthalten.

Die Auszeichnungssprache SGML erlaubte erstmals, als integralen Teil des Dokuments strukturierte Informationen in ein definiertes Element des Dokuments, den Kopfteil („header“), aufzunehmen, mit denen andere Elemente des Dokuments beschrieben und damit erschlossen werden. Diese strukturierten Informationen können von Programmen ausgelesen und weiterverwendet werden.

In der praktischen Anwendung können im Kopfteil von HTML-Dokumenten in „Meta-tags“ Informationen über die Inhalte im Rumpfteil („body“) als Werte der Attribute „description“ und „keywords“ angegeben werden. „Description“ kann eine unstrukturierte, kurze Inhaltsbeschreibung des Dokuments enthalten, in „keywords“ können Stichwörter als Indexterme angegeben werden (Ferber 2002:930). Problematisch hierbei ist, dass keinerlei Qualitätskontrolle stattfindet. Im Gegenteil, diese Mechanismen werden in der Praxis auch häufig dazu benutzt, um Suchmaschinen in die Irre zu führen, also einen Dokumenteninhalte vorzugaukeln, der in Wahrheit nicht vorhanden ist, um einen günstigen Platz in der Ergebnispriorisierung der Suchmaschinen zu belegen. Deshalb vertrauen die Entwickler vieler Suchmaschinen heute nicht mehr Metadaten, die von Webseiten-Betreibern selbst zur Verfügung gestellt werden. Google zum Beispiel ignoriert die Inhalte der „Meta-tags“ sowohl für die Gewichtung als auch sogar für die Indizierung. Man kann beklagen, dass dieser Effekt aus dem eigentlich sachgerechten, einfach zu handhabenden „Meta-tags“-Mechanismus entstanden ist²⁵, aber im anarchischen, den Marktmechanismen unterworfenen Internet ist er nicht zu vermeiden, da die Betreiber kommerzieller Webseiten in harter Konkurrenz zueinander stehen und mit allen Mitteln um möglichst hohe Wahrnehmungsraten kämpfen. Gegenmaßnahmen könnten sein, dass Metadaten geprüft und verifiziert werden durch das im Entstehen begriffene System von „Trust“-Diensten und digitalen Signaturen, wie es vom World-Wide-Web-Konsortium geplant ist.²⁶

Dieses Problem der Validität von HTML-Metadaten liegt also in der Organisationsform des Internets begründet und könnte auch bei anderen technisch-konzeptionellen Metadatenstrukturen entstehen. Abgesehen davon besteht bei HTML aber noch ein weiteres Problem, das durch den begrenzten Funktionsumfang dieser Auszeichnungssprache hervorgerufen wird: Mit den relativ beschränkten Mitteln von HTML kann keine semantische Annotation der Metadaten vorgenommen werden, was aber zu Erschließungszwecken unabdingbar ist. Denn Erschließung soll die gezielte Suche nach Dokumenten ermöglichen. Ohne semantische Strukturierung kann beispielsweise bei der Suche nach einem Namen nicht unterschieden werden, ob Dokumente von oder über eine Person dieses Namens gesucht und gefunden werden.

²⁵ So stellt Priscilla Caplan fest: „It is unfortunate that misuse of metadata by an unethical minority can prevent the effective use of metadata by the majority.“ (Caplan 2003:46)

²⁶ Berners-Lee, Tim: Semantic Web Road map. 1998.

<http://www.w3.org/DesignIssues/Semantic.html> (Stand vom 12. 05. 2004); siehe auch Caplan 2003:47 und Niederée 2003:195

XML dagegen erlaubt im Rahmen der anerkannten Dokumententypen, bestimmte Elemente von Dokumenten als mit einer bestimmten Bedeutung versehen zu definieren und so genannte semantische „tags“ zu verwenden. Dadurch kann zum Beispiel innerhalb der Struktur des Dokuments gekennzeichnet werden, dass ein Name die Bedeutung „Autor des Dokuments“ hat, wodurch sowohl menschliche LeserInnen als auch automatisierte Auswertungsprogramme der Suchmaschinen mit diesem Namen dieselbe Bedeutung verknüpfen können. Mit XML ist also die erschließende Beschreibung von Internetdokumenten mittels semantisch strukturierter Metadaten möglich geworden (Mühlhäuser 2002:875), darüber hinaus wird XML auch für geeignet gehalten, „...semiotische Thesauri in einer völlig neuen Konstruktion aufzubauen“ (Umstätter 2000:249). Um hier Interoperabilität zu erreichen, muss allerdings zunächst definiert werden, welche Bedeutung welchen „tags“ zukommen soll. Zu diesem Zweck wird beim W3-Konsortium ein neuer Standard zur Ressourcenbeschreibung entwickelt: Resource Description Framework (RDF), der ein Meilenstein auf dem Weg zum „Semantic Web“ sein soll.

2.4.3. „Semantic Web“

„Semantic Web“ sollte nicht als inhaltsarme modische Wortschöpfung missverstanden werden; das Schlagwort benennt einen der aktuell wichtigsten Trends in der Weiterentwicklung des World Wide Web (Niederée 2003:193). Die Zielsetzung oder „Vision“ – wie es mit einem ebenfalls gern missverstandenen Modewort oft ausgedrückt wird – seiner Protagonisten im World-Wide-Web-Konsortium ist es, durch semantische Anreicherung von Information eine neue Qualität in der Bereitstellung von Inhalten und Diensten im Internet zu erreichen, wodurch die Kooperation zwischen Menschen und Computern erleichtert werden soll (Niederée 2003:193 f.). Die Kooperation zwischen Mensch und Maschine kann zwar kein Selbstzweck sein, letztlich kann die Zielsetzung nur in der besseren Nutzung der im Internets vorhandenen Informationen, z.B. von maschinell erzeugten Suchergebnissen, durch Menschen liegen. Aber erstens heißt es im englischen Original nicht ganz so eindimensional „better enabling computers and people to work in cooperation“ (Berners-Lee/Hendler/Lassila 2001), und zweitens werden noch viel ehrgeizigere Ziele artikuliert, beispielsweise: „The goal of the semantic web is to express real life.“²⁷

Im einzelnen soll dies durch folgende Weiterentwicklungen der technischen Konzepte erreicht werden:

Während Metadaten in HTML Teil des Dokuments selbst sind (Werte im „header“), wird mit dem Resource Description Framework (RDF) eine von der konkreten Syntax eines bestimmten Dateiformats unabhängige Metasyntax definiert (Dörr/Enderle/Hauße 2004:405). RDF ist grundsätzlich syntaxunabhängig konzipiert, soll aber nach einem Vorschlag des World-Wide-Web-Konsortiums ptimalerweise in XML-Anwendungen umgesetzt werden.

Die Beschreibung von HTML-Ressourcen mittels „Meta-tags“ erlaubt nicht die Behandlung komplexerer Zusammenhänge in den Metadaten, da

²⁷ Berners-Lee: What the Semantic Web can represent. 1998.
<http://www.w3.org/DesignIssues/RDFnot.html> (Stand vom 12. 05. 2004)

Relationen zwischen den Elementen nicht ausgedrückt werden können. Das Resource Description Framework soll diesen Mangel beheben. Die Grundidee des „Semantic Web“ geht noch einen Schritt weiter: Auch die Relationen zwischen Web-Objekten selbst sollen semantisch und kontextabhängig beschrieben werden können (Miller/Hillmann 2002:57)²⁸.

Mit RDF wird also ein umfassender Beschreibungsstandard für Metadaten und mit XML ein erweiterter und flexiblerer Standard zur Strukturierung von Dokumenten im WWW definiert (Ferber 2002:930). Modelle und Formate für eine semantische Annotation von Informationsobjekten durch Metadaten sind also zumindest als standardisierte Konzepte vorhanden. Um Internetressourcen nun aber auch tatsächlich mit semantisch bedeutsamen Metadaten versehen zu können, fehlen noch zwei wichtige Komponenten:

- Methoden und Software-Werkzeuge zur tatsächlichen Durchführung der Annotation. Angesichts der geschätzten drei Milliarden Dokumente, aus denen das World Wide Web zur Zeit besteht (Niederée 2003:194)²⁹, muss dieser Vorgang offensichtlich zumindest teilweise automatisiert werden. Fortgeschrittene Verfahren der automatischen Indexierung könnten dafür tauglich sein, befinden sich aber teilweise auch noch im Entwicklungsstadium (siehe unten Kapitel 3.3.1.).
- Um die Inhalte der Metadaten mit interpersonell valider Bedeutung zu versehen, müssen in den jeweiligen Wissens- oder Fachgebieten (in der Sprache des W3C: Domänen) so genannte Ontologien entwickelt werden.

Der Terminus „Ontologie“ wird in diesem Zusammenhang nicht im philosophischen Sinn als die Lehre vom Sein oder vom Seienden verstanden, sondern so wie er in der Informatik und Informationswissenschaft im Zusammenhang mit Forschungen zur Künstlichen Intelligenz enger definiert wurde. Er wird dort allgemein nach Gruber definiert als Spezifikation einer Konzeptionalisierung³⁰. Eine Ontologie soll also die Konzepte einer Domäne, eines Wissensgebiets, erfassen und zueinander in Beziehung setzen (Niederée 2003:195), es handelt sich damit um ein formal definiertes System von Konzepten und ihren Relationen. Ontologien unterscheiden sich von anderen Begriffssystemen durch die Möglichkeit, auch Relationen über Relationen auszudrücken, was im Resource Description Framework als „reification“ bezeichnet wird (Niederée 2003:97). Lässt man diese Eigenschaft von Ontologien weg, dann bleibt letztlich ein Konzept übrig, was im bibliothekarisch-dokumentarischen Bereich als „kontrolliertes Vokabular“ mit den Ausprägungen Klassifikation und Thesaurus bezeichnet wird.³¹

Im bibliothekarisch-dokumentarischen Bereich sind solcherart Konzepte selbstverständlich nicht unbekannt, schließlich beschäftigt man sich dort schon seit Jahrtausenden mit der Erschließung und Klassifikation des Weltwissens³².

²⁸ „... and provide mechanisms and standards to support the semantics or context of how objects in the Web relate to one another.“

²⁹ Empirisch belegt sind Zahlen zum Umfang und Wachstum des World Wide Web in den Jahren 1998 bis 2002: Im Juni 2002 existierten 3.080.000 Webseiten („Web sites“) mit 1,4 Milliarden Dokumenten („Web pages“) im öffentlichen Teil des Internets, siehe O’Neill/Lavoie/Bennett 2003

³⁰ „An ontology is a specification of a conceptualization.“ Gruber 1993

³¹ siehe Caplan: „Ontologies ... are ... somewhat similar to thesauri“ (Caplan 2003:52)

³² vgl. z.B. zur historischen Entwicklung von Thesauri: Schwarz/Umstätter 1999:197 ff.

Eric Miller³³ und Diane Hillmann stellen fest: „These technologies ... are based on concepts the library community takes for granted. It is interesting that these kinds of technologies, which are becoming mainstream, are actually the bread and butter of what we do in libraries.” (Miller/Hillmann 2002:57)

Die weitere Entwicklung des „Semantic Web“ bleibt abzuwarten, gerade im Bereich der Ontologien befinden sich die Arbeiten des World-Wide-Web-Konsortiums noch in einem recht frühen Stadium (Niederée 2003:197). Allerdings wird die Entwicklung vom World-Wide-Web-Konsortium stetig vorangetrieben, inzwischen sind Empfehlungen zur Anwendung einer „Web Ontology Language“ erschienen,³⁴ die die dezentrale Entwicklung von Ontologien in allen Wissensgebieten ermöglichen sollen. Auch eine Open-Source-Software für Ontologie-Entwicklungsumgebungen ist schon vorhanden, sie wurde von den Universitäten Stanford und Manchester in einem gemeinsamen Projekt³⁵ entwickelt (Horridge 2004).

Trotzdem ist es noch fraglich, ob und in welchem Zeitraum es gelingen wird, die kritische Masse von semantisch annotierten Internetressourcen zu erreichen, die notwendig wäre, damit der Standard sich von der Theorie zur Praxis entwickelt. Es ist sicher nicht damit zu rechnen, dass sich das „Semantic Web“ in kurzer Zeit als Paradigma für das gesamte Web durchsetzt, eher könnten „viele kleinere Semantic Webs aus gezielten Initiativen entstehen“ (Niederée 2003:198). Trotz der extremen Schnellebigkeit der Materie scheint diese Einschätzung aus dem vergangenen Jahr auch heute noch Bestand zu haben.

2.5. Dublin-Core-Metadatenstandard

2.5.1. Beschreibung, Definition

Mit der Entwicklung des Dublin-Core-Metadatenstandards³⁶ wurde das Metadatenkonzept schon frühzeitig von der Bibliothekswelt rezipiert (Dörr/Enderle/Hauffe 2004:404). 1995 wurde bei einem Workshop in Dublin/Ohio, dem Sitz des einflussreichen Online Computer Library Center (OCLC), erstmals ein internationaler und interdisziplinärer Metadatenstandard definiert. Beteiligt waren 52 VertreterInnen von Bibliotheken und Archiven, aus der sozial- und naturwissenschaftlichen Forschung und aus dem Bereich des sich gerade ausbreitenden öffentlichen Internets.³⁷ Zu diesem Zeitpunkt war die Entwicklung des World Wide Web noch lange nicht so weit fortgeschritten wie dies aus

³³ Eric Miller war als Mitarbeiter von OCLC einer der Gründer der Dublin-Core-Metadaten-Initiative und ist heute in leitender Funktion an der Entwicklung des „Semantic Web“ beteiligt, siehe <http://www.w3.org/People/EM/> (Stand vom 12. 05. 2004)

³⁴ offizielle „recommendation“ des W3C vom 10. Februar 2004, siehe <http://www.w3.org/TR/2004/REC-owl-features-20040210/> (Stand vom 12. 05. 2004)

³⁵ CO-ODE – A Collaborative Open Ontology Development Environment

³⁶ Standards sind „jedwede in einem Konsensprozess entstandenen Spezifikationen“, wie es das DIN-SICT-Papier „Strategie für die Standardisierung der Informations- und Kommunikationstechnik“ definiert, vgl. Krause/Niggemann/Schwänzl 2003:19

³⁷ „Fifty-two librarians, archivists, humanities scholars and geographers, as well as standards makers in the Internet, Z39.50 und Standard Generalized Markup Language (SGML) communities, met to identify the scope of the problem, to achieve consensus on a list of metadata elements...“ (Weibel 1995)

heutiger Perspektive vielleicht anzunehmen wäre, sie steckte vielmehr noch in den Kinderschuhen. Die frühe Dublin-Core-Initiative diene damit der wechselseitigen Befruchtung zwischen den unterschiedlichen Welten des Web und des BID-Bereichs (Caplan 2003:2).³⁸

Der Dublin-Core-Metadatenstandard besteht aus einem Minimalsatz von 15 Erschließungselementen.³⁹ Er sollte ursprünglich dazu dienen, die Probleme der bibliothekarischen Erschließung von Internetressourcen zu lösen, indem er den AutorInnen thematischer Websites die Möglichkeit an die Hand gab, auf einfache Weise Metadaten selbst zu erstellen.

Mit folgenden Prinzipien wurde versucht, größtmögliche Flexibilität für fachlich, formal und inhaltlich unterschiedliche Anwendungen zu erreichen:

- jedes Element ist optional, kann also auch fehlen;
- jedes Element kann beliebig oft wiederholt werden („every element is optional and repeatable“, Baker 2000).

Zudem sind auch die Regeln, die bestimmen, wie die 15 Metadatenelemente gefüllt werden sollen, gewollt unscharf, so dass von „fifteen fuzzy buckets“ gesprochen wird (Baker 2000). Denn wenn die Regel präziser wären, würden Anwender unweigerlich in Versuchung kommen, sie zu umgehen oder zu verletzen (ebd.)⁴⁰. Wegen dieser Flexibilität und wenig festgelegten Regelmäßigkeit und auch wegen seiner simplen Syntax wird der Metadatenstandard auch als „Pidgin for Digital Tourists“ (ebd.) bezeichnet, was die Zielsetzung des Konzepts – Austausch rudimentärer Informationen, mithin Kommunikation über Sprach- oder andere Grenzen hinweg – einprägsam verdeutlicht.

Die angestrebte Modularität und Universalität des Metadatenstandards, also seine Eignung für inhaltlich höchst unterschiedliche Informationsobjekte in unterschiedlichen fachlichen Zusammenhängen, wird durch die Legosteine-Analogie anschaulich beschrieben:

„In a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way. Thus, application designers should be able to benefit from significant re-usability as they gather existing modules of metadata and 'snap' them together as much as individual Lego blocks can be assembled into larger structures. ... Children think nothing of mixing cowboy themes and pirate themes and undersea exploration themes. While the 'semantics' of such combinations may not always be obvious to adults, children don't seem to be bothered by such incongruities. Similar flexibility should be achievable in the metadata architecture of the Web.“

³⁸ „The organizers of the first Dublin Core workshop were active participants in the World Wide Web Consortium (W3C), an (at that time) infant organization concerned with managing the development of the equally infant Web. As such, the early Dublin Core initiative served as an agent of cross-fertilization between the library and Web communities.“

³⁹ siehe <http://www.dublincore.org/documents/dces/> (Stand vom 12. 05. 2004); die offizielle deutsche Übersetzung des Dublin-Core-Satzes findet sich hier: <http://www.mpib-berlin.mpg.de/DOK/metatagd.htm> (Stand vom 12. 05. 2004)

⁴⁰ „If the rules of Dublin Core were more precise, people would inevitably bend them.“ (Baker 2000)

(Duval/Hodgins/Sutton/Weibel 2002)

In jährlichen Konferenzen wurde der Standard weiterentwickelt. Die 15 Elemente sind zwar seit 1996 stabil, aber eine entscheidende inhaltliche Erweiterung war die Einführung von „Qualifiers“ (Element Refinements and Encoding Schemes). In ihrer Form als „refinements“ ermöglichen sie Differenzierungen innerhalb einzelner Elemente, also eine hierarchische Struktur. So kann zum Beispiel im Element „Datum“ genauer differenziert werden, um welche Daten es sich handelt: Erstellungsdatum, Veröffentlichungsdatum, Datum der letzten Änderung etc. Außerdem können DC-Qualifizierungen in ihrer Ausprägung als „encoding scheme“ auf Kontextinformation oder allgemein anerkannte Interpretationsschemata für einzelne Feldinhalte Bezug nehmen, beispielsweise auf Klassifikationen wie Dewey Decimal Classification oder Fachthesauri, deren Benutzung ebenfalls innerhalb des Dublin-Core-Elements angezeigt wird (Baker 2000).

Diese Entwicklung wird in der Fachöffentlichkeit als Annäherung an die bibliothekarisch-dokumentarischen Erschließungsmethoden eingeschätzt: „... das Metadatenkonzept, das sich zunächst als die bessere bzw. einfachere Alternative zur bibliothekarischen Erschließung von Internetressourcen verstand, [holt] in gewisser Weise Professionalisierungsstadien nach, welche das auf einem bibliothekarischen Regelwerk und Datenformat beruhende Modell bereits seit Jahrzehnten durchlaufen hatte.“ (Dörr/Enderle/Hauße 2004:405)

Seit Januar 2003 ist Dublin Core als internationaler Metadatenstandard anerkannt (Dekkers/Weibel 2003).⁴¹ Die Dublin-Core-Metadateninitiative ist inzwischen auch wahrhaft international geworden, es gehören ihr heute VertreterInnen von 50 Staaten an, Übersetzungen des Metadatenstandards in mehr als 20 Sprachen liegen vor.⁴²

Die Dublin-Core-Metadateninitiative empfiehlt offiziell die Implementierung des Standards mithilfe von RDF/XML. Schon 1999 stellte die DC-Arbeitsgruppe fest: „The Dublin Core Data Model Working Group agrees that RDF provides adequate flexibility and functionality to fully support encoding DC metadata. ... The Dublin Core Data Model applies RDF as an enabling mechanism, providing consistent guidelines to support Dublin Core applications.“ (Bearman et al. 1999) Inzwischen sind auch Richtlinien zur Implementierung von Dublin Core in XML veröffentlicht worden.⁴³

Der Dublin-Core-Metadatenstandard ist damit technisch gesehen eine Referenzanwendung im Rahmen der Entwicklung von RDF, inhaltlich gesehen ein Versuch, im extrem heterogenen, unstrukturierten Internet eine Insel der Systematisierung („island of control“, Ferber 2002:931) aufzubauen. Damit soll die im World Wide Web „verloren gegangene Konsistenz aus der Welt der

⁴¹ ISO 46/SC 4 N515; ISO 15836:2003(E): Information and documentation - The Dublin Core metadata element set. 2003.

⁴² siehe <http://dublincore.org/resources/translations/> (Stand vom 12. 05. 2004); siehe auch Nagamori/Baker/et.al., die schon 2001 über „A Multilingual Metadata Schema Registry Based on RDF Schema“ berichten

⁴³ „Guidelines for implementing Dublin Core in XML“, April 2003, siehe <http://dublincore.org/documents/dc-xml-guidelines/> (Stand vom 12. 05. 2004)

Bibliotheken und Fachdatenbanken teilweise neu etabliert werden“ (Krause/Niggemann/Schwänzl 2003:26).

Auch Michael Gorman, ein hochrangiger Funktionär der American Library Association, der dem Konzept von Metadaten sehr kritisch gegenübersteht und Dublin Core als „ill-formulated subset of the MARC record“ (Gorman 2002:181 f.) bezeichnet, muss zugestehen, dass das Metadatenkonzept einen dritten Weg ermöglicht, sozusagen die goldene Mitte zwischen der teuren, aber effektiven intellektuellen Erschließung und kostengünstiger Volltextindizierung, die den NutzerInnen aber nur ineffektive Stichwortsuche bietet.⁴⁴ Und so war es von Anfang an auch gemeint, schon im ersten darüber publizierten Aufsatz beschrieb Stuart Weibel die beiden Extreme möglicher Erschließungsmethoden für digitale Dokumente: Suchmaschinen-Index und bibliothekarische Katalogisierung. Weibel sah im gerade entwickelten Metadatenkonzept das Potential für eine alternative Lösung, die zwischen diesen Extreme vermitteln könnte (Weibel 1995).⁴⁵

2.5.2. Anwendungsbeispiele

Anwendungsbeispiele des Dublin-Core-Metadatenstandards existieren international vor allem in drei Bereichen:

- in Retrodigitalisierungsprojekten;
- bei Wissenschaftsportalen oder „subject gateways“;
- als kleinster gemeinsamer Nenner für komplexere Metadatenstrukturen, der eine Suche über mehrere Datenquellen ermöglicht.

(Caplan 2003: 85).

Für die letztgenannte Anwendung als „kleinster gemeinsamer Nenner“ ist als bekanntestes Beispiel die Open Archives Initiative (OAI)⁴⁶ anzuführen, die als Minimalanforderung benennt, dass alle teilnehmenden Internetressourcen in der Lage sind, Daten im unqualifizierten Dublin-Core-Standard zu exportieren (Caplan 2003:85). Die 1999 ins Leben gerufene Initiative hat es sich zur Aufgabe gemacht, eine offene Struktur für den Metadatenaustausch zu schaffen, die für alle Anbieter wissenschaftlicher und wissenschaftlich relevanter Dokumente offen ist. Neben dem unvermeidlichen OCLC sind daran viele namhafte US-amerikanische und angelsächsische Universitäten und Forschungseinrichtungen beteiligt, auch eine Vertreterin des deutschen Max-Planck-Instituts hat einen Sitz im Steuerungskomitee. Gefördert wird die Initiative unter anderem von der National Science Foundation, dem US-Gegenpart der Deutschen Forschungsgemeinschaft.

Als herausragendes Beispiel eines schon weit fortgeschrittenen Retrodigitalisierungsprojekts sei das Colorado Digitization Program⁴⁷ genannt, in dem seit 1998 Sammlungen von Museen, Archiven, Bibliotheken aus den Staaten des US-amerikanischen Westens digitalisiert und damit für Forschungszwecke

⁴⁴ „The idea behind metadata ist that there is some Third Way, approximately halfway between cataloguing (expensive and effective) and keyword searching (cheap and ineffective).“ (Gorman 2002:180 f.)

⁴⁵ „An alternative solution that promises to mediate these extremes...“

⁴⁶ <http://www.openarchives.org/> (Stand vom 12. 05. 2004)

⁴⁷ <http://www.cdphheritage.org/> (Stand vom 12. 05. 2004), siehe auch Garrison 2001.

und auch für die interessierte Öffentlichkeit zugänglich gemacht werden⁴⁸. Es ist in den USA ein Pilotprojekt für die Zusammenarbeit von Bibliotheken, Museen und Archiven und dient als Referenzprojekt für die Anwendung und Weiterentwicklung des Dublin-Core-Metadatenstandards im Rahmen von Projekten retrospektiver Digitalisierung (Garrison 2001). Interessant für die jüngsten Entwicklungen im Retrodigitalisierungsbereich ist ein aktuelles Projekt der British Library, in dem eine historische Zeitung („The Penny Illustrated Paper“) großen Umfangs – 40.000 Seiten, 500.000 Abbildungen – digitalisiert wird mithilfe einer von OCLC entwickelten Software, die sowohl Texterkennung als auch Metadatenvergabe weitgehend zu automatisieren verspricht.⁴⁹

Zusammen mit den USA nimmt Deutschland eine Vorreiterrolle bei der Entwicklung und Anwendung des Metadatenstandards ein. Schon 1998 existierten über 30 Projekte in Deutschland, die Dublin Core anwendeten (Baker 1998). Im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten META-LIB-Projekts wurde im Zeitraum von 1997 bis 2002 einerseits die theoretische Weiterentwicklung des DC-Standards betrieben – der 7. Dublin-Core-Metadaten-Workshop wurde 1999 in Frankfurt ausgerichtet. Andererseits wurde auch die Umsetzung des Metadatenstandards in praktische Anwendungen befördert. META-LIB hatte zum Ziel, deutschlandweite Richtlinien für den Einsatz des Dublin-Core-Standards zur Erschließung digitaler und digitalisierter Ressourcen in wissenschaftlichen Bibliotheken zu entwickeln. Beteiligt waren unter anderem die Deutsche Bibliothek und die Staats- und Universitätsbibliothek Göttingen. Wichtige Projektergebnisse sind die Handreichungen „METADISS-Format des Datensatzes für Online-Hochschulschriften Der Deutschen Bibliothek und des Projekts Dissertationen Online“ und „Empfehlungen zur Definition eines Metadaten-Core-Sets für die verteilte Suche über die Virtuellen Fachbibliotheken“⁵⁰ (Oehlschläger 2003:180). Daraufhin wurden von den einzelnen Virtuellen Fachbibliotheken so genannte „Application Profiles“ entwickelt, mit denen die spezifischen Anforderungen des jeweiligen Fachgebiets abgebildet werden konnten. Ein „Application Profile“ kann Elemente verschiedener Schemata oder „Namespaces“ enthalten, zum Beispiel von unqualifiziertem und qualifiziertem Dublin-Core-Standard, und diese darüber hinaus auch mit anderen „Namespaces“, beispielsweise aus dem Learning-Object-Metadaten-Standard (LOM, siehe Kamke/Zimmermann 2003:346 f.) kombinieren. Durch die Einrichtung von „Registries“ wurde zwischen den diversen „Application Profiles“ trotz unterschiedlicher Ausprägungen des Dublin-Core-Standards Interoperabilität gewährleistet (Fischer 2003:199 f., Caplan 2003:39 ff.).

Die Virtuellen Fachbibliotheken haben sich inzwischen mit den großen Informationsverbänden und mit anderen Institutionen zum Projekt „Vascoda“⁵¹ zusammengeschlossen, in dem ein umfassendes, interdisziplinäres Internetportal für wissenschaftliche Information in Deutschland entsteht. Seit Mitte 2003 steht es

⁴⁸ „Through digitization, the Collaborative Digitization Program, a collaboration of archives, historical societies, libraries and museums of the West, enables access to cultural, historical and scientific heritage collections, thereby increasing understanding of the past and informing future generations.“ <http://www.cdphheritage.org/about/mission.html> (Stand vom 12. 05. 2004)

⁴⁹ vgl. Jordan 2004: „... software, which incorporates optical character recognition and automatic XML tagging for improved searching...“

⁵⁰ http://www2.sub.uni-goettingen.de/metacore/empfehlungen/core_set.pdf (Stand vom 12. 05. 2004)

⁵¹ <http://www.vascoda.de> (Stand vom 12. 05. 2004)

der Öffentlichkeit zur Verfügung; zur Zeit besteht es größtenteils noch aus einer Metasuchmaschine, die gleichzeitige Suchabfragen über die diversen dort miteinander integrierten Datensammlungen erlaubt. Ein „Application Profile“ zur Definition des Vascoda-spezifischen Metadatenstandards auf Grundlage von Dublin Core ist in der Entwicklungsphase.

In den „Empfehlungen des Wissenschaftsrats zur digitalen Informationsversorgung durch Hochschulbibliotheken“ vom 16. Juli 2001 wurde konstatiert: „Der bereits weit verbreitete Dublin Core ist als Mindeststandard für die Erschließung digitaler Objekte anzusehen und sollte ... allgemein übernommen werden.“⁵² Einzelne Universitätsbibliotheken wie Essen und Jena haben schon seit einigen Jahren digitale Bibliotheken unter Verwendungen des Dublin-Core-Metadatenstandards aufgebaut. Das erste Projekt war Miles, der multimediale Lehr- und Lernserver Essen⁵³, der seit 1998 als Kooperationsprojekt von Rechenzentrum, Bibliothek und Medienzentrum auf der Plattform eines IBM-Contentmanagementsystems Inhalte von Lehrveranstaltungen und später auch Dissertationen beherbergt und publiziert.

Dieses Projekt ist also wie die Nachfolgeprojekte in Jena, Münster, Leipzig, Göttingen etc.⁵⁴ inhaltlich eine interne digitale bzw. multimediale Universitätsbibliothek, die Anwendung des Dublin-Core-Metadatenstandards hat sich also entfernt von ihrem ursprünglichen Ausgangspunkt, der Erschließung von Ressourcen im öffentlichen Internet. Aber auch in diesen Anwendungsgebieten hat sich der Metadatenstandard bewährt, weil er optimal geeignet ist für die zu erschließenden heterogenen Dokumente. Je nach Anwendung und Zielsetzung werden die Dokumente in Dokumenttypen differenziert, deren Erschließung unterschiedlich geregelt ist und trotzdem noch konform geht mit dem Dublin-Core-Standard. So werden zum Beispiel in der University Multimedia Electronic Library (UrMEL) an der Thüringer Universitäts- und Landesbibliothek Jena alle vorkommenden Dokumenttypen wie z.B. Dissertationen, Audio- und Videomaterialien, Studienapparate etc. in drei Kategorien eingeteilt (Kuhles/Wefers 2001:223 f.), bei denen unterschiedliche Anforderungen an die Archivierungsintensität und an die Erschließungstiefe gestellt werden, die zu verschiedenen Ausprägungen in den einzelnen Metadatenelementen führen können, ohne dass der Standard infrage gestellt wird. Daran zeigt sich die hohe Flexibilität und breite Einsetzbarkeit des Dublin-Core-Metadatenstandards.

⁵² <http://www.wissenschaftsrat.de/texte/4935-01.pdf> (Stand vom 12. 05. 2004)

⁵³ <http://miless.uni-essen.de/about/> (Stand vom 12. 05. 2004)

⁵⁴ siehe die Zusammenstellung auf <http://miless.uni-essen.de/about/nachnutzung.xml> (Stand vom 12. 05. 2004)

2.5.3. Einschätzung der zukünftigen Entwicklung

Kann ein Metadatenstandard wie Dublin Core die klassische Erschließung mithilfe der traditionellen Regelwerke AACR oder RAK/RWSK zukünftig ersetzen oder ergänzen? Diese oft gestellte Frage muss im Rahmen dieser Arbeit unbeantwortet bleiben. Für die Erschließung analoger Informationsobjekte könnten die klassischen Regelwerke, seien sie nun US-amerikanisch-angelsächsischer Herkunft oder im deutschsprachigen Raum entwickelt, möglicherweise noch lange angemessene und pragmatisch nützliche Erschließungsinstrumente bleiben. Über zukünftige Entwicklungen in diesem Bereich können im Rahmen dieser Arbeit keine Aussagen getroffen werden.

Was digitale Dokumente angeht, ist die Sachlage jedoch anders. Für die Erschließung digitaler Informationsobjekte waren die klassischen Regelwerke noch nie als angemessen oder geeignet zu bezeichnen, weil sie vom Ansatz her auf die Erschließung analoger Medien ausgelegt sind. Ein Metadatenstandard wie Dublin Core ist speziell auf diesen Bedarf hin entwickelt worden. Fraglich ist, ob gerade der Dublin-Core-Metadatenstandard sich durchsetzen wird, oder ob er in Konkurrenz zu anderen Standards oder Methoden stehen könnte, die bessere Chancen haben, eine kritische Masse von Anwendungen zu erreichen. Wie oben gezeigt, sind die Initiativen rund um das „Semantic Web“ keine Konkurrenz, sondern eher als ein technisch-konzeptioneller Rahmen für eine metadatenbasierte Erschließung von Internetressourcen zu werten, deshalb ist keine Entwicklung konkurrierender Standards ersichtlich.

Schließlich ist für die zukünftige Durchsetzung des Dublin-Core-Metadatenstandards ins Feld zu führen, dass er unter Federführung von OCLC entwickelt wurde und von dieser mit großer Marktmacht ausgestatteten Institution (siehe oben Kapitel 2.1.) weiterhin unterstützt wird. Zudem sind schon zahlreiche Anwendungen des Dublin-Core-Metadatenstandards vorhanden, einige sind in Deutschland schon weit über das Projekt- und Experimentierstadium hinaus im produktiven Betrieb erfolgreich.

Nicht ganz so klar abzusehen ist dagegen zur Zeit die weitere Entwicklung innerhalb der vielfältigen Metadatenprojekte: Welche Lösungsansätze werden für das Problem gefunden, dass trotz Metadatenstandard die Erschließungsdaten in unterschiedlichen Dublin-Core-Anwendungen notwendigerweise – durch den flexiblen Standard bedingt – heterogen sind? Sally McCallum, die in leitender Funktion im Bereich Format- und Strukturentwicklung von Metadaten bei der Library of Congress tätig ist, formuliert es so: „The web and XML developments have spawned a great deal of diversity that could be useful and destructive at the same time. ... Has DC got staying power for cross-domain interoperability?“ (McCallum 2003:183)

Einerseits scheint es global-zentralistische Tendenzen zu geben, zum Beispiel die Idee eines Internationalen „Authority File“: Library of Congress, OCLC und die Deutsche Bibliothek wollen in einem gemeinsamen Projekt exemplarisch für den Bereich der Personennormdaten ein „Virtual International Authority File“ aufbauen, um die Realisierbarkeit des Ansatzes einer globalen Normdatei nachzuweisen (Hengel 2003:4 ff.). Wegen seiner hochkarätigen Teilnehmerschaft aus der deutschsprachigen und der US-

amerikanischen Fachwelt wird dieses Projekt als Meilenstein in der internationalen Kooperation von Bibliotheken bezeichnet.⁵⁵ Auch im LEAF-Projekt⁵⁶, in dem 15 BID-Institutionen aus zehn europäischen Staaten von der Staatsbibliothek Berlin koordiniert zusammenarbeiten, wird auf europäischer Ebene versucht, die diversen Normdateien zu einer größeren Struktur verbindlicher Vokabularien oder Normdateien zu vereinigen.

Angesichts der globalen Struktur des Internets, die mit den Mitteln heutiger Nationalstaaten oder auch supranationaler Organisationen sicherlich auch in den nächsten Jahrzehnten weiterhin unregulierbar, heterogen und letztlich anarchisch oder nur den Marktgesetzen unterworfen bleiben wird, erscheint es unwahrscheinlich, dass diese von einem zentralistischen Ansatz ausgehenden Projekte Erfolg versprechend sein könnten. Dies wird in der Fachwelt heute auch überwiegend so gesehen. So wird festgestellt, dass heute kaum noch jemand der Vorstellung nachhänge, „der Dokumentenraum ließe sich durch globale Standardisierungsabsprachen über alle Teilbereiche hinweg homogenisieren, sich organisatorisch wieder auf einige wenige Mitspieler reduzieren oder über ein hierarchisch organisiertes Modell der Kooperation gestalten. Ganz im Gegenteil, die heutigen Vorstellungen gehen von einer noch stärkeren Dezentralisierung bei der Dokumenterstellung, -erschließung und -verteilung aus, wodurch die ‚anarchischen Tendenzen‘ weiter zunehmen.“ (Krause/Niggemann/Schwänzl 2003:27) Deshalb muss der „verbleibenden und unvermeidlichen Heterogenität“ (ebd.) durch angemessene Strategien begegnet werden, z.B. durch die Entwicklung von Transfermodulen, so genannten Crosswalks und Crosskonkordanzen. Auf diesem Gebiet werden aller Voraussicht nach in naher Zukunft die nächsten Entwicklungsschritte zu erwarten sein.

Ein weiterer zukunftsweisender Ansatz scheint die Integration von Metadatenkonzepten mit Suchmaschinentechnologie zu sein. Optimistisch wird in der Literatur festgestellt, dass Suchmaschinenentwickler und Metadatenspezialisten beginnen, die Vorteile des jeweils anderen Systems anzuerkennen und immer mehr Funktionalitäten der „Gegenseite“ zu übernehmen (Younger 2002:8)⁵⁷. Realisierte Anwendungsbeispiele dieser Entwicklung sind zwar noch nicht in Sicht, aber im Rahmen des CARMEN⁵⁸-Projekts wird an einer Verbindung verschiedener Metadaten schemata „in einem spezifischen, qualitativ hochwertigen Index“ gearbeitet, der letztlich „zum Einsatz der Metadaten in Suchmaschinen im bibliothekarischen und fachwissenschaftlichen Bereich“ (Tappenbeck/Wessel 2001:569) führen soll.

Umgekehrt wird auch versucht, bibliothekarische Daten in kommerzielle Suchmaschinen zu integrieren. In einem gemeinsamen Pilotprojekt von Google und OCLC soll eine Teilmenge der im „WorldCat“ aggregierten weltweiten Opac-

⁵⁵ „This is a landmark in international library cooperation“, sagte der Präsident von OCLC am 5. Februar 2004 auf der 7. Internationalen Bielefeld Konferenz (Jordan 2004)

⁵⁶ das Akronym LEAF steht für „Linking and Exploring Authority files“

⁵⁷ „...that the developers of richly structured databases such as library catalogs and the developers of Web search engines are recognizing the benefits of their counterparts and are increasingly adopting each other's features.“

⁵⁸ das Akronym setzt sich zusammen aus „Content Analysis, Retrieval and Metadata: Effective Net-working“ und ist seit 1999 ein Projekt im Rahmen von Global Info, siehe Tappenbeck/Wessel 2001:566 ff.

Daten aus Bibliothekskatalogen der kommerziellen Suchmaschinendatenbank zur Verfügung gestellt werden, was letztlich dazu führen könnte, dass der geschmähte Google-Nutzer dann doch in seine heimliche Bibliothek geht: „An information seeker who starts a search using Google could end up finding the items needed in a nearby library.“ (Jordan 2004) Dieses Pilotprojekt nutzt zwar nicht den Dublin-Core-Metadatenstandard, sollte aber aus Gründen der Vollständigkeit in diesem Ausblick nicht fehlen.

Die Integration von Metadatenkonzepten und so genannten „Web Services“ scheint dagegen noch Zukunftsmusik zu sein. Hinter dem Schlagwort „Web Services“ stehen Konzepte, mit denen versucht wird, die einzelnen Komponenten des „Semantic Web“ durch Automatisierung zu effektivieren. Ob allerdings im „Semantic Web“ für Metadaten je ein Szenario erreicht wird, wie es Jay Jordan im folgenden beschreibt, erscheint fraglich:

„The one machine says to the other, ‚Please send me a Dewey number for this record.‘ The other machine says, ‚here’s your Dewey number.‘ While it’s a bit more complicated than that, it is an example of the kind of services that we will be seeing more of as the semantic Web becomes reality.“

(Jordan 2004)

3. Praxisbeispiel: Dokumentenmanagement bei Greenpeace Deutschland

3.1. Hintergrund: Wissensmanagement

Die Umweltschutzorganisation Greenpeace ist in besonders hohem Maße auf das Wissen ihrer MitarbeiterInnen angewiesen, um erfolgreich agieren zu können. Sie kann als Musterbeispiel einer „wissensbasierten“ Organisation gelten. Übereinstimmend wird in der Literatur zum Querschnittsthema Wissensgesellschaft, Wissensmanagement etc. festgestellt, dass beim aktuellen Stand der ökonomischen Entwicklung in den Industrieländern der Faktor Wissen quasi zur Produktivkraft geworden ist, dass „intellektuelles Kapital“ entstanden ist, dessen Generierung und effektiver Einsatz für Unternehmen, Organisationen, Staaten und Gesellschaften zu entscheidenden Erfolgsfaktoren geworden sind (Willke 2001:1 ff., Haun 2002:5 ff.). Eine Organisation wie Greenpeace besitzt keine anderen „Produktionsmittel“⁵⁹ als das Wissen ihrer MitarbeiterInnen – einerseits Faktenwissen über Umweltthemen, zum Beispiel Messergebnisse über die Toxizität bestimmter Chemikalien, andererseits Wissen über erfolgreiches oder auch weniger erfolgreiches Agieren der Organisation in der Vergangenheit, über Kampagnenmethoden, über Reaktionen in der Öffentlichkeit, über erfolgreiche oder weniger erfolgreiche Methoden des Fundraisings etc.

Je nach der Art des Wissens sind unterschiedliche Methoden des Wissensmanagements angemessen. Grundlegend wird nach Nonaka und

⁵⁹ Produktionsmittel in Anführungszeichen, weil eine gemeinnützige Nichtregierungsorganisation nichts Geldwertes produziert. Deshalb sind auch Ansätze aus der klassischen Betriebswirtschaftslehre, die sich mit dem „Return on Investment“ für Wissensmanagement beschäftigen, auf den Non-Profit-Bereich schwer übertragbar.

Takeuchi, den „Erfindern“ des Wissensmanagements, Wissen unterschieden in explizites und implizites Wissen. Explizit wäre beispielsweise in Tabellen niedergelegtes Wissen über Messergebnisse toxischer Chemikalien, ein Beispiel für implizites Wissen wäre die Fähigkeit einer erfahrenen Pressesprecherin, intuitiv die „richtigen“ Worte für die jeweilige Zielgruppe zu finden. Es liegt auf der Hand, dass gerade letztere Form von Wissen besonders schwer zu vermitteln ist und dies sicherlich nicht nur über seine Externalisierung geleistet werden kann, also indem eine Wissensträgerin versucht, ihr Wissen zu artikulieren, zu verschriftlichen und auf diesem Wege anderen Organisationsmitgliedern zu vermitteln. Dieser Prozess wird wahrscheinlich in großen Teilen viel erfolgversprechender mit den Methoden der Sozialisation durchgeführt, also indem zum Beispiel neue MitarbeiterInnen durch gemeinsame Handlungspraxis wie Schüler von Meistern lernen (Willke 2001:14). Verschriftlichte Informationsobjekte wären in diesem Prozess der Sozialisation nur von untergeordneter Wichtigkeit, während sie für explizites Wissen und seine Vermittlung die geeignete Speicher- und Vermittlungsform darstellen. Denn „per Definition ist .. jegliches vom Rechner verarbeitbare Wissen explizit“ (Staab 2002:200). Das macht deutlich, dass ein Dokumentenmanagementsystem nur einen bestimmten Teil von Wissensmanagement abdecken kann, und dass genau geprüft werden muss, für welche Teile des in der Organisation vorhandenen, zu erwerbenden und zu vermittelnden Wissens welche Methoden des Wissensmanagements eingesetzt werden sollten.

Eine wissensbasierte Organisation kann nur dann zur Generierung von innovativem Wissen gelangen, „wenn sie die schwierigen und voraussetzungsreichen Übergänge zwischen explizitem und implizitem Wissen in routinisierte organisationale Prozesse fasst, die fördern, dass individuelles Wissen artikuliert und durch Zugänglichkeit verbreitet wird“ (Willke 2001:15). Um dieses hehre Ziel zu erreichen, unter dem sich die vielfältigen Maßnahmen von Wissensmanagement letztlich subsumieren lassen, ist gezielte Planung notwendig. Und je kritischer die Ressource Wissen für die jeweilige Organisation ist, umso größer ist die Bedeutung eines professionellen Wissensmanagements (Willke 2001:86).

3.2. Dokumentenmanagement: Ausgangssituation, Problembeschreibung

Das Dokumentenmanagementsystem kann von seinen Funktionen und Inhalten her als Teil eines Wissensmanagementsystems kategorisiert werden, denn es ist ein Speichersystem für Dokumente, in denen explizites, verschriftlichtes Wissen der OrganisationsmitarbeiterInnen aufbewahrt wird. Zudem ermöglicht es die Kooperation der MitarbeiterInnen durch seine Kommunikationsfunktionen (siehe Haun 2002:210).

Seit 1997 existiert bei Greenpeace Deutschland ein internes Dokumentenmanagementsystem. Auf der Basis der Software Lotus Notes, die sich in den neueren Versionen Lotus Domino nennt, wird sowohl elektronische Kommunikation (E-Mail, Newsgroups, Gruppenkalender) als auch die interne Ablage digitaler Dokumente betrieben. Die Software Lotus Notes existiert seit 1989, sie ist seit 1995 ein Produkt von IBM und wird als

Groupware bezeichnet, weil sie sowohl Funktionen des Dokumentenmanagements als auch eines Kommunikationssystems bietet und damit die Übergänge zwischen elektronischer Kommunikation einerseits, Dokumentenbearbeitung und -ablage andererseits reibungslos ermöglicht. Alle Informationsobjekte, seien es E-Mails oder andere Dokumente, werden in Datenbanken gespeichert. Diese haben ein proprietäres Datenbankformat, das weder relational noch dateibasiert ist. Die Datenbanken sind volltextindiziert.

Zudem bietet die Software eine Entwicklungsplattform, d.h. auf der Basis vordefinierter Elemente können spezifische Anwendungen entwickelt werden. Software-Entwicklungswerkzeuge existieren auf verschiedenen Ebenen, von einfachen, auch NutzerInnen zugänglichen Makro-ähnlichen Strukturen, z.B. zur Erstellung von Agenten, bis hin zu einer objektorientierten Skriptsprache inklusive Klassenbibliotheken. Außerdem wird die Software-Nutzung an unterschiedlichen Standorten oder mit mobiler Anbindung optimal unterstützt, da verschiedene Stände einer Datenbank über so genannte Replikationen leicht miteinander abgeglichen werden können.

Lotus Notes / Domino wird häufig von Großunternehmen im Dienstleistungssektor, wie Banken und Versicherungen, eingesetzt, aber auch vereinzelt im Bibliotheksbereich⁶⁰. Greenpeace Deutschland benutzt Lotus Notes / Domino für

- elektronische Kommunikation: Alle BenutzerInnen verfügen über eine eigene Maildatenbank, es gibt einige gemeinsame Maildatenbanken mit generischen Adressen wie z.B. „mail@greenpeace.de“ für spezifische Zwecke. Außerdem existiert ein System von Newsgroups zur Kommunikation mit anderen Greenpeace-Büros auf internationaler Ebene und deutschlandweit mit den ca. 70 Gruppen von ehrenamtlichen Greenpeace-MitarbeiterInnen.
- Dokumentenmanagement im Rahmen von selbstentwickelten Datenbankstrukturen.

Der Terminus „Dokumentenmanagement“ ist an dieser Stelle möglicherweise missverständlich, weil er im betriebswirtschaftlichen Bereich oft für Systeme anderer Art verwendet wird. Im kommerziellen Bereich kommen häufig Systeme zum Einsatz, in denen ursprünglich analoge Dokumente wie eingehende Aufträge, Rechnungen etc. zunächst gescannt und dann per automatischer Texterkennung (OCR) in ein Datenbanksystem integriert werden, um dann im Rahmen eines definierten Work-flow-Prozesses intern weiterbearbeitet zu werden. Im Gegensatz dazu ist die überwiegende Mehrheit der Dokumente im Greenpeace-Dokumentenmanagementsystem ursprünglich digital, d.h. sie kommen entweder schon in digitaler Form ins System, wie zum Beispiel Presseagenturmeldungen per Satellitenübertragung, oder sie werden von Greenpeace-MitarbeiterInnen digital erstellt. Es existieren keine definierten Abläufe in Form von Work-flows oder ähnlichem, die die Erstellung, Bearbeitung und Ablage von Dokumenten strukturieren.

⁶⁰ vgl. die Beschreibung eines Lotus-Notes-basierten Systems zur automatischen Indexierung retrodigitalisierter Bibliotheksbestände bei Hauer 2003:194

Zur Zeit benutzen etwa 250 MitarbeiterInnen 31 so genannte Bereichs- und Themendatenbanken, die einerseits parallel zur Organisationsstruktur⁶¹ pro Bereich oder Team eingerichtet wurden, andererseits für bereichsübergreifende Arbeitsgruppen oder Gremien. Die Zugriffsrechte auf diese Datenbanken sind ebenfalls entlang den Organisationseinheiten vergeben, im Regelfall hat ein Bereich/Team oder eine Arbeitsgruppe Zugriffsrechte auf „seine“ Datenbank und kann diese bei Bedarf erweitern auf MitarbeiterInnen anderer Bereiche. Außerdem existieren sechs Datenbanken mit speziellen Funktionen, zum Beispiel „Schwarzes Brett“, Meldungen von Presseagenturen, eigene Presseerklärungen und andere Veröffentlichungen von Greenpeace Deutschland. Auf diese Datenbanken haben meist alle MitarbeiterInnen Lesezugriff, die Schreibzugriffe sind der jeweiligen Fachabteilung, z.B. dem Personal- oder Kommunikationsbereich, vorbehalten. Zur Zeit sind in den Bereichs- und Themendatenbanken etwa 70.000 Dokumente vorhanden, es gehen ca. 100.000 Agenturmeldungen jährlich ein.

In jeder Datenbank gibt es bereichsspezifische Schlagworte, Kategorien, Dokumentenarten etc., die zur intellektuellen Erschließung der Dokumente dienen.⁶² Außer dem Titel-Feld sind alle Erschließungsfelder optional, müssen also nicht belegt werden. Mehrfachwerte sind in jedem Feld möglich. Die Kategorisierung bzw. Verschlagwortung der Dokumente wird zwar zum Zeitpunkt der Dokumentenbearbeitung mit kontrolliertem Vokabular durchgeführt, dieses Vokabular kann aber von den NutzerInnen leicht und flexibel innerhalb jeder Datenbank geändert und erweitert werden und ist deshalb nicht als kontrolliertes Vokabular im klassisch dokumentarischen Sinne zu bezeichnen (vgl. Knorz 1997:125; Burkart 1997:164). Zudem existiert kein organisationsweiter Thesaurus, kein Klassifikationssystem oder ähnliches. Es handelt sich also um ein dezentrales, nicht hierarchisches Erschließungssystem, das sehr flexibel ist und von der Konzeption her den Bedürfnissen der einzelnen Themenbereiche oder Arbeitsgruppen optimal entsprechen soll.

Darüber hinaus sind alle Datenbanken per Volltextindex erschlossen. Diese Volltextindizierung ist eine Grundfunktionalität der Software Lotus Notes und steht damit überall im System zur Verfügung, beispielsweise auch in den Mail-Datenbanken der BenutzerInnen. Das führt zu recht guter Bedienbarkeit, weil dieselben Funktionalitäten zur Suche in allen zur Verfügung stehenden Datenbanken genutzt werden können und die BenutzerInnen nur eine Abfragesprachsyntax und eine Suchmethode erlernen müssen.

Für die Suchabfragen können die übliche Boole'schen Operatoren und Platzhalterzeichen benutzt werden⁶³, darüber hinaus ist eine so genannte unscharfe Suche möglich, die grammatikalische Varianz und Schreibfehler abfangen soll.⁶⁴ Mit diesen Suchabfragen kann einerseits im Volltext der

⁶¹ siehe Anhang 1: Organigramm Greenpeace Deutschland e.V.

⁶² siehe Anhang 2: Maske eines Dokuments in einer Bereichsdatenbank mit den Erschließungsfeldern

⁶³ siehe „Domino R5 Supported Syntax für Full Text Searching“: http://www-1.ibm.com/support/docview.wss?rs=0&q1=domino&q2=7003210&uid=swg27003210&loc=en_US&cs=utf-8&cc=us&lang=en (Stand vom 12. 05. 2004)

⁶⁴ Auf welchem Indizierungsverfahren dies basiert, ist leider nicht zweifelsfrei zu recherchieren, da wie allgemein üblich die Details der hinter dem Datenbanksystem

Dokumente gesucht werden, andererseits können auch gezielt Feldinhalte ausgewertet werden, zum Beispiel die Felder, die zur intellektuellen Erschließung dienen. Durch die Kombination dieser Methoden stehen damit vielfältige Möglichkeiten zur Verfügung, die Dokumentensammlungen nach Stichworten zu durchsuchen.⁶⁵

Nach Jahren der Benutzung zeigen sich nun folgende Probleme⁶⁶:

- Die dezentrale Lösung führt dazu, dass „dasselbe“ Dokument unbemerkt in mehreren Bereichsdatenbanken vorhanden sein kann, eventuell sogar in unterschiedlichen Versionen.
- Die Datenbanken werden in den Organisationseinheiten sehr unterschiedlich genutzt, Dokumentensammlungen sehr unterschiedlicher Qualität und Konsistenz sind entstanden.
- Auch die zur Erschließung vorgesehenen Felder werden sehr unterschiedlich genutzt, die Systematik der Erschließungsbegriffe ist für bereichsfremde MitarbeiterInnen meist nicht selbsterklärend und damit nicht nutzbar.
- Es existiert keine Archivierungsmöglichkeit.
- Das interne Dokumentenmanagementsystem hat keine technische oder logische Verbindung zu den beiden später entstandenen Systemen:
 - Contentmanagementsystem für die eigene Website (www.greenpeace.de)
 - von Greenpeace International betriebenes Intranetsystem.

Zu fragen ist nun, ob und inwiefern Überschneidungen dieser Problembereiche mit den oben dargestellten aktuellen Themen in der bibliothekarisch-dokumentarischen Fachdiskussion zur Erschließung digitaler Dokumente festgestellt werden können und ob möglicherweise dort diskutierte Lösungsstrategien auch für die Situation bei Greenpeace Deutschland in Frage kommen könnten.

3.3. Überschneidungsbereiche mit allgemeiner Erschließungsdiskussion und -praxis

In diesem Abschnitt soll versucht werden, Ansatzpunkte für Problemlösungsstrategien herauszuarbeiten, indem Parallelen zwischen der beschriebenen Situation im Praxisbeispiel und der allgemeinen Erschließungssituation digitaler Dokumente im BID-Bereich herausgearbeitet werden. Folgende Themenbereiche könnten die Situation im Praxisbeispiel verallgemeinernd charakterisieren:

liegenden informationstechnischen Methoden und Verfahren von der Herstellerfirma nicht preisgegeben werden. Lässt man das übliche Marketing-Wortgeklänge außer Betracht, kann zumindest darauf geschlossen werden, dass Grundformenreduktion („stemming“) im Indizierungsverfahren enthalten ist und ein statistisches Verfahren zugrunde liegt.

⁶⁵ s. Anhang 3: Suchfenster in einer Bereichsdatenbank

⁶⁶ Diese Sammlung der Problempunkte ist das Ergebnis von Vorarbeiten und Abstimmungen im Rahmen des Greenpeace-Projekts „Studie zur Optimierung des Dokumentenmanagements“, siehe Anhang 4: Projektstartdokument.

- Problem der Ressourcen-Identifikation:
„Dasselbe“ Dokument könnte unbemerkt an mehreren Stellen im System vorhanden sein, möglicherweise in verschiedenen Versionen. Das führt zur Frage nach der Identifikation „eines“ Dokuments: Was soll als ein eindeutig bezeichnetes Dokument im Greenpeace-System gelten? Hier könnte eine Parallele zur Situation im öffentlichen Internet vorliegen, auch dort gibt es Bemühungen, Dokumente eindeutig zu bezeichnen, das World-Wide-Web-Konsortium hat als Lösungsansatz so genannte Uniform Resource Identifier (URI)⁶⁷ vorgeschlagen. Teilweise ist dieses Thema als Unterfall des Persistenzproblems im öffentlichen Internet (siehe oben Kapitel 2.3.) zu bewerten, es ist allerdings genauso wie dieses im Bereich des öffentlichen Internets auch nicht ansatzweise in der Praxis gelöst. Zudem besteht hier meiner Ansicht nach gerade keine Parallele zwischen der Situation im öffentlichen Internet und im Greenpeace-Intranet.

Denn beim Dokumentenmanagementsystem von Greenpeace Deutschland handelt es sich um ein zwar heterogenes, aber immer noch überschaubares Intranet-System, das zentral administriert wird und deshalb nicht denselben strukturellen Problemen wie das öffentliche Internet ausgesetzt ist. In einer Intranet-Umgebung wären – anders als im öffentlichen Internet – zentralistische Lösungsansätze durchaus als erfolgversprechend einzuschätzen, sie sind für diverse andere Anwendungsbereiche auch bereits software-technisch realisiert, man denke zum Beispiel an Redaktionssysteme oder Softwareentwicklungsumgebungen, in denen eine klare, eineindeutige Identifikation von funktionalen Objekten (Artikel bei Redaktionssystemen, Programmcode bei Softwareentwicklungsumgebungen) Voraussetzung für die Konsistenz des ganzen Systems ist und deshalb selbstverständlich gewährleistet wird. Zudem bieten solche Systeme die Möglichkeiten einer Versionskontrolle, d.h. mehrere Ausprägungen eines eindeutigen Objekts können vorliegen, die meist verschiedene Stadien der Bearbeitung repräsentieren. In der bei Greenpeace realisierten Lotus-Notes-Datenbankumgebung werden eineindeutige so genannte Dokumenten-IDs automatisch generiert, die allerdings keine Funktionalität wie Versionskontrolle etc. beinhalten, sondern lediglich zur technischen Identifikation der Dokumente zum Beispiel für Replikationszwecke dienen. Falls Bedarf nach Work-flow-artigen Funktionalitäten besteht, könnte diese Dokument-ID als Grundlage für Anwendungsentwicklung dienen.

Damit erweist sich das Grundproblem der Ressourcen-Identifikation zwar auf der Erscheinungsebene im öffentlichen Internet und in den Greenpeace-Intranet-Datenbanken als ähnlich gelagert, durch völlig unterschiedliche Rahmenbedingungen sind aber keine Parallelen in möglichen Lösungsansätzen erkennbar.

- Heterogenität der Dokumentensammlungen:
Durch die Strukturierung der Dokumente in voneinander unabhängige Themen- und Bereichsdatenbanken sind Dokumentensammlungen entstanden, die von sehr unterschiedlicher Konsistenz und Qualität gekennzeichnet sind, also ein hohes Maß an Heterogenität aufweisen. Dieses Problem ist im öffentlichen Internet genauso oder vielmehr in weit größerem Umfang vorhanden. Die Frage

⁶⁷ <http://www.w3.org/Addressing/Activity> (Stand vom 12. 05. 2004)

des sachgerechten Umgangs mit Heterogenität stellt sich also gleichermaßen, lässt sich in dieser Allgemeinheit allerdings nicht sinnvoll weiterverfolgen. Es muss zunächst geklärt werden, worin das Problem liegt, das durch Heterogenität hervorgerufen wird. Oben konnte gezeigt werden, dass im öffentlichen Internet das Problem darin besteht, valide Fachinformation zu finden inmitten all der vielfältigen, heterogenen Informationsobjekte, die eben gerade keine valide Fachinformation darstellen oder nur am Rande relevant sind. In einem internen Dokumentenmanagementsystem dagegen ist die Qualität und Relevanz eines Informationsobjekts noch viel weniger eindeutig zu bestimmen, da sie situationsabhängig ist und letztlich nur von den NutzerInnen selbst beurteilt werden kann: Was dazu dient, den Informationsbedarf in einer konkreten Arbeitssituation zu decken, oder was beim Erreichen eines bestimmten Arbeitsergebnisses nützt, wird NutzerInnen eines Wissensmanagementsystems als qualitativ angemessenes Informationsobjekt erscheinen und ist es damit auch. Objektive Maßstäbe für Relevanz und Qualität eines Informationsobjekts lassen sich in diesem Bereich kaum festlegen, da sie abhängig sind von den Zielen, die damit erreicht werden sollen. Deshalb ist kein Lösungsansatz für das Problem der Heterogenität der Greenpeace-Dokumentensammlungen aus der allgemeinen Fachdiskussion über Heterogenität im öffentlichen Internet ersichtlich.

- **Fehlende Archivierungsmöglichkeit:**

Die Problematik rund um das Thema Langzeitarchivierung wurde im Rahmen dieser Arbeit zwar in ihrer Ausprägung als Problem der technischen Formate ausgeklammert, sie darf aber im Praxisbeispiel in ihrer inhaltlichen Dimension nicht vernachlässigt werden. Für Institutionen im BID-Bereich ist dieser Aspekt der Archivierungsthematik so selbstverständlich, dass er sozusagen impliziter Teil ihrer Tätigkeit ist. Bibliotheken sind tendenziell auf die Ewigkeit ausgelegt, Archive haben gerade die Funktion, Kulturgut zu bewahren. Für eine relativ junge Organisation wie Greenpeace⁶⁸ dagegen ist hier noch einiges an Problembewusstsein und Problemlösungskompetenz aufzuholen. Wissensmanagement hat unter anderem auch die Funktion, ein „Organisationsgedächtnis“ zu schaffen. Die Funktionen eines solchen „Gedächtnisses“ sind allerdings andere als in den Institutionen des BID-Bereichs, denn sie orientieren sich immer an den Zielen der jeweiligen Organisation, während Bibliotheken der allgemeinen oder wissenschaftlichen Öffentlichkeit zur Verfügung stehen und damit keine spezielle Zielsetzung mit ihrer Archivierungstätigkeit verfolgen, sondern Vollständigkeit im Rahmen ihres Sammelauftrags anstreben müssen. Lösungsansätze zum Problem der fehlenden Archivierungsmöglichkeit im Praxisbeispiel lassen sich also eher im Bereich des Wissensmanagements als im BID-Bereich finden.

- **Probleme der Erschließung:**

Die Dokumentensammlungen sind sowohl durch Volltextindizierung als auch durch intellektuell vergebene Begriffe, also Metadaten, erschlossen. Die Ergebnisse der intellektuellen Erschließung werden als problematisch dargestellt, da sie ebenso wie die Dokumentensammlungen selbst sehr heterogen sind und teilweise dazu führen, dass die Kategorisierungssysteme nicht selbsterklärend

⁶⁸ Greenpeace wurde in Kanada gegründet. Das deutsche Büro wurde im November 1980 in Hamburg gegründet, siehe <http://www.greenpeace.org/deutschland/?page=/deutschland/greenpeace/ueberuns/index#gruendung> (Stand vom 12. 05. 2004)

sind, also zumindest für bestimmte Benutzergruppen keine erschließende Funktion mehr haben.

Daraus lassen sich folgende Fragen ableiten:

- Ist die Volltextindizierung optimierbar? Gerade im Bereich digitaler Bibliotheken werden Fragen der automatischen Indizierung wieder vertieft diskutiert, deshalb erscheint es lohnend, dieser Frage weiter nachzugehen.
- Metadaten sind vorhanden, wenn auch inkonsistent und in unterschiedlicher Qualität. Da liegt die Frage nahe: Könnte ein Standard nützen? Wie weit wäre der Weg zum Dublin-Core-Metadatenstandard?

Insgesamt scheint die Erschließungssituation Parallelen aufzuweisen zu der Thematik, die oben im Rahmen der „Einschätzung der zukünftigen Entwicklung“ diskutiert wurde. Denn wie bei digitalen Informationsobjekten, die mithilfe von Dublin-Core-Metadaten erschlossen und zudem volltextindiziert sind, kommen folgende, technisch möglicherweise realisierbare Weiterentwicklungen auch beim Dokumentenmanagement von Greenpeace Deutschland in Frage:

- automatisierte Metadatenerzeugung
- Integration von Metadatenerschließung und Suchmaschinentechnologie.

Eine automatisierte Metadatenerzeugung wäre technisch-funktional gesehen ein Unterfall des Vorschlags, die Volltextindizierung zu optimieren, sollte also in diesem Zusammenhang diskutiert werden. Zur näheren Diskussion wurden damit drei Bereiche isoliert, in denen Parallelen zwischen Praxisbeispiel und allgemeiner Fachdiskussion gefunden wurden:

- Optimierung der Volltextindizierung
- Anwendung des Dublin-Core-Metadatenstandards
- Integration von Metadatenerschließung und Suchmaschinentechnologie.

3.3.1. Optimierung der Volltextindizierung

Grundsätzlich besteht die Herausforderung bei der Volltextindizierung von unstrukturierten Dokumenten darin, dass anders als bei einer Suche in strukturierten Datenbanken mithilfe einer Abfragesprache wie SQL meist nicht präzise Suchbegriffe oder Suchabfragen formuliert werden können, sondern ein unscharfer Informationsbedarf von NutzerInnen möglichst optimal befriedigt werden muss, auch wenn die Anfragen nur unvollständig oder vage formuliert werden können (Ferber 2002:916). Das liegt nicht etwa am Unvermögen der BenutzerInnen, dem beispielsweise mit Schulungsmaßnahmen begegnet werden könnte, sondern zum großen Teil in der Natur der Sache selbst: Anlass der Suche ist ein Wissensdefizit, das auch die Suche selbst beeinflusst, da die NutzerInnen in dieser Situation meist nicht in der Lage sind, eine genaue und präzise Frage über ein zu diesem Zeitpunkt unklares Problem zu stellen. Das Paradoxon besteht darin, dass die NutzerInnen eine genaue Beschreibung des Problems erst nach einer erfolgreichen Suche und nach Auswertung der gefundenen Informationen formulieren könnten (Nohr 2003:18). Deshalb müssen auf Volltextindizierung basierende Suchsysteme mit unscharfen Anfragen⁶⁹ umgehen können, um optimale Ergebnisse zu erreichen. Andererseits muss die Menge der gefundenen

⁶⁹ zur Unschärfe in Datenbank- und Retrievalsystemen vgl. ausführlich Fühles-Ubach 1997. <http://www.ib.hu-berlin.de/%7Ewumsta/ubach/index.htm> (Stand vom 12. 05. 2004)

Dokumente sinnvoll begrenzt und in eine Rangfolge gebracht werden, um den NutzerInnen ein handhabbares Ergebnis zu liefern, es muss also eine Gewichtung der Suchergebnisse stattfinden.

Eine objektive, allseits akzeptierte Systematisierung der Verfahren der Volltextindizierung ist zur Zeit kaum möglich (Nohr 2003:31), da die meisten in Forschung und Praxis realisierten Indexierungssysteme Elemente mehrerer Verfahren kombinieren. Von der Funktionsweise her lassen sich folgende Verfahren unterscheiden:

- das elementare Verfahren der einfachen Stichwortextraktion, oft auch als Freitext- oder Volltextinvertierung bezeichnet. Hier wird das Problem der sprachlichen Vielfalt ohne weitere Optimierung gänzlich auf den Retrievalvorgang und damit auf die NutzerInnen verlagert (Knorz 1997:130; Nohr 2003:29), unscharfe Anfragen sind nicht möglich.
- statistische Verfahren, die von philologischen Erkenntnissen abgeleitete statistische Gesetzmäßigkeiten der jeweiligen Sprache nutzen, um Indexterme auszuwählen und zu gewichten (Nohr 2003:33 ff.). Diese Verfahren wurden inzwischen zum Vektorraummodell weiterentwickelt, mit dessen Hilfe Ähnlichkeitsmaße für Dokumente und Anfragen bestimmt werden können, die Grundlage für Rangfolgen in Suchergebnislisten sein können (Ferber 2002:923).
- computerlinguistische Verfahren, deren einfachstes die Grundformenreduktion ist, wo also als Indexterm nicht die im Dokument vorhandene grammatikalische Form eines Wortes verwendet wird, sondern seine Grundform. Die computerlinguistischen Verfahren können algorithmisch (regelbasiert) sein oder auf elektronischen Wörterbüchern beruhen (Nohr 2003:30, Ferber 2002:919).

Die meisten zur Zeit in handelsüblichen Datenbank- und Informationssystemen integrierten Volltextindizes beruhen auf einer Kombination von statistischen und linguistischen Indizierungsverfahren. Die Syntax der Abfragesprache in diesen Systemen bietet meist die Möglichkeit, Boole'sche Operatoren und Platzhalterzeichen („wild cards“) zu nutzen, also zum Beispiel durch Platzhalterzeichen am Wortende eine Trunkierung vorzunehmen und damit schon einen Großteil der Probleme grammatikalischer Varianz zu umgehen.

Diese Situation liegt auch bei den volltextindizierten Lotus-Notes-Datenbanken des Greenpeace-Dokumentenmanagementsystems vor. Wie oben schon erwähnt, kann die genaue Indizierungstechnologie nicht im Detail recherchiert werden, weil die Firma IBM dies als Geschäftsgeheimnis behandelt. Die Suchmaschine wird als Global Text Retrieval (GTR) bezeichnet und ist sprachenunabhängig, das heißt es wird nur ein Index für Dokumente in allen möglichen Sprachen und Zeichensätzen erstellt, die Suchabfragen können ebenfalls in einer beliebigen Sprache gestellt werden. Anhänge an die Dokumente werden ebenfalls indiziert. Über die üblichen Suchabfrage-Funktionalitäten wie Boole'sche Operatoren hinaus wird in den letzten Versionen von Lotus Domino versucht, die Abfragesprache-Syntax den von Internet-Suchmaschinen gewohnten Modalitäten anzupassen, so wird zum Beispiel die Eingabe mehrerer Suchbegriffe ohne verbindenden Operator als „near“-Suche interpretiert, das heißt es wird nach dem Vorkommen beider Suchbegriffe (entspricht einer Und-Verknüpfung) in einem zu gewichtenden Bereich von Nähe innerhalb eines Dokuments gesucht. Außerdem, und das ist

recht ungewöhnlich für ein Volltextindexsystem, können auch komplexe Suchabfragen auf die Inhalte jedes Datenbankfeldes formuliert werden. Das bedeutet, dass auch die für intellektuelle Erschließung genutzten Felder gezielt für Suchen genutzt werden können, bei Bedarf auch gleichzeitig mit einer unspezifischen Suche nach dem Auftreten bestimmter Suchbegriffe irgendwo im Dokument.

Es handelt sich also schon in seiner jetzigen Ausprägung um ein recht leistungsfähiges Indizierungssystem. Einige mögliche Funktionalitäten werden nicht genutzt und böten damit Raum für Optimierungsmöglichkeiten:

- Durch die Einrichtung eines dedizierten Servers für eine so genannte „Domänensuche“ könnte ein Index über alle Datenbanken und auch noch über andere Ressourcen wie zum Beispiel ein Dateisystem geschaffen werden. Damit könnten mit einer Suchabfrage alle Datenbanken und anderen Ressourcen durchsucht werden.
- Durch das Zusatzprodukt „Discovery Server“ würden Möglichkeiten automatischer Indizierung bzw. Klassifizierung geboten.

Allgemein können beim automatisierten Indexieren mit dem Ziel einer Klassifikation der Dokumente, häufig anhand eines festgelegten Vokabulars, zum Beispiel einem Fachthesaurus, folgende Verfahren unterschieden werden:

- Mustererkennungsverfahren, mit denen Terme oder Termgruppen innerhalb eines Textes erkannt werden. Diese Verfahren haben häufig eine „selbstlernende“ Komponente und werden mit Systemen kombiniert, die durch das Indexierungsverhalten menschlicher Indexierer lernen, Dokumente in Klassifikationssystemen einzuordnen (Nohr 2003:75 ff.).
- Begriffsorientierte Verfahren: Diese beruhen nicht mehr auf einer Extraktion vorhandener Textterme, sondern abstrahieren von der gegebenen Wortwahl vorliegender Dokumente auf die Bedeutung der Texte und verwenden dabei Methoden der Textwort- bzw. Textanalyse. Sie sind bisher nur für begrenzte und homogene Dokumentensammlungen zu kleinen Themengebieten realisiert und auch für diese Anwendungen extrem aufwändig, deshalb bisher noch von sehr geringer praktischer Relevanz (Nohr 2003:79 ff.).

Diese Verfahren automatisierter Indexierung erlauben NutzerInnen auch das Blättern („Browsen“) durch ein Klassifikationssystem, was einen erheblichen Mehrwert im Vergleich zu den klassischen Indexierungsverfahren darstellt, aber auch großen Aufwand bei der Erstellung und Pflege des zugrundeliegenden Klassifikationsschemas erfordert. Sie stellen also eine Möglichkeit dar, automatisiert Metadaten zu erzeugen. Der oben erwähnte „Discovery Server“ der Firma IBM wäre eine technische Umsetzungsmöglichkeit, aber auch Produkte anderer Firmen wären mit dem jetzigen Dokumentenmanagementsystem bei Greenpeace Deutschland kompatibel.

Als Beispiel einer Anwendung im bibliothekarischen Bereich sei ein Projekt an der Vorarlberger Landesbibliothek Bregenz genannt: Digitalisierung von Aufsätzen und anderen Texten mit maschineller Inhaltserschließung (Hauer 2003:192 ff.). Ebenfalls auf der Basis eines Lotus-Notes-Systems werden dort Inhaltsverzeichnisse von Monographien und Aufsätze aus Fachzeitschriften

eingescannt, per automatischer Texterkennung in Volltext umgewandelt und dann mit einem zusätzlichen Programmmodul CAI-Engine (Computer Aided Indexing, kein Lotus- oder IBM-Produkt, sondern ein so genanntes „third party“-Produkt) automatisiert indexiert, wobei Thesauri und Klassifikationen hinterlegt sind und die Indexierung steuern. Wenn im Indexat kein erkennbares Äquivalent zu einem Thesaurusbegriff gefunden wird, wird ein „als wichtig erkannter Term mit seinem Termgewicht“ (Hauer 2003:195) ausgegeben, was dann zu manueller Weiterbearbeitung führt. Auch für die automatisch erzeugten Indexbegriffe findet eine Qualitätskontrolle durch fachkompetente BibliotheksmitarbeiterInnen statt.

Fraglich ist, ob diese Methoden geeignet wären, die Erschließung der Dokumente in den Datenbanken bei Greenpeace Deutschland zu verbessern. Denn Voraussetzung wäre, dass ein organisationsweites Klassifikationsschema oder ein Fachthesaurus entwickelt würde. Ob und mit welchem Aufwand dies bei der Heterogenität der behandelten Themen und auch bei den unterschiedlichen Arbeitsmethoden (wissenschaftliche Arbeit auf der einen Seite, Aktionsorientierung auf der anderen) möglich wäre, kann an dieser Stelle nicht eingeschätzt werden. In jedem Fall kann diese Frage nicht auf der theoretischen Ebene entschieden werden, sondern es müsste zunächst ein konkreter Bedarf auf Seiten der NutzerInnen festgestellt und genauer spezifiziert werden.

3.3.2. Dokumentenerschließung durch standardisierte Metadaten

Wie oben gezeigt, wird als ein großes Problem der Greenpeace-Dokumentensammlungen die Heterogenität und unterschiedliche Qualität der Metadaten gewertet. Die Anwendung des Metadatenstandards Dublin Core könnte eine konsistentere Erschließung über die unterschiedlichen Datenbanken hinweg bewirken, hätte aber auch noch weitergehende positive Effekte: Sie könnte die Zukunftssicherheit der Erschließung verbessern und auch die Kompatibilität mit anderen Datenhaltungssystemen. So wird zur Zeit bei der Dachorganisation Greenpeace International ein „Digital Asset Management System“ projektiert, in dem ebenfalls über die Verwendung des Dublin-Core-Metadatenstandards nachgedacht wird. Es ist offensichtlich, dass die Kompatibilität der beiden Systeme auch über Systemgrenzen hinweg durch einen gemeinsamen Metadatenstandard signifikant erhöht würde.

Wie dargestellt, sind Felder für Metadaten in den Bereichs- und Themendatenbanken enthalten. Ihre Funktionalität entspricht auch den Grundsätzen des Dublin-Core-Standards, sie sind optional⁷⁰ und wiederholbar. Es ist also zu fragen: Was würde noch fehlen für eine Erschließung, die mit dem Dublin-Core-Metadatenstandard konform ist?

Letztlich würde die Beantwortung dieser Frage in der Entwicklung eines „Application Profiles“ für einen Greenpeace-Metadatenstandard auf Grundlage von Dublin Core münden. Im Rahmen dieser Arbeit ist dies aber nicht zu leisten, daran müssten unter Projektbedingungen VertreterInnen diverser NutzerInnengruppen mitarbeiten und ihre Sachkunde einbringen. Man denke nur an die Entwicklung von kontrolliertem Vokabular – ein Standard-

⁷⁰ Außer dem Titel-Feld muss keines der Erschließungsfelder belegt sein, damit ein Dokument abgespeichert werden kann, siehe oben 3.1.

Klassifikationsschema wie Dewey Decimal Classification wird höchstwahrscheinlich nicht in Frage kommen, also müsste ein spezifischer Fachthesaurus oder ein Klassifikationsschema mit ähnlicher Funktion entwickelt werden. Das erfordert auf der einen Seite, dass die Sachkunde beispielsweise von CampaignerInnen in ihrem jeweiligen Fachgebiet eingebracht wird; auf der anderen Seite müssen sich die künftigen NutzerInnen in einem strukturierten Prozess auf die jeweiligen Begriffe einigen und einen Konsens über Relationen, Zusammenhänge und die Details ihrer künftigen Benutzung finden.⁷¹

Im folgenden werden deshalb im ersten Schritt die einzelnen Metadaten-Elemente lediglich kurz benannt und erläutert⁷², auf ihre Relevanz in der Greenpeace-Umgebung überprüft und festgestellt, ob sie in den bestehenden Datenbanken⁷³ schon vorhanden sind:

1. Titel: relevant, vorhanden.
2. Verfasser oder Urheber: die Person/en oder Organisation/en, die den intellektuellen Inhalt verantworten.
Ein Feld „Autor“ ist in den Greenpeace-Datenbanken vorhanden, es wird bei Erstellung des Dokuments automatisch mit dem Namen des aktuellen Systembenutzers gefüllt. Dies muss nicht in allen Fällen die Person sein, die für den intellektuellen Inhalt des Dokuments verantwortlich ist, oftmals ist die Ablage von Dokumenten in der Datenbank Aufgabe der Assistenzen, die Dokumenteninhalte wurden aber von Campaignern, PressesprecherInnen etc. erstellt. In solchen Fällen könnte ein Verfasserfeld also nützlich sein. Das Metadaten-Element ist also relevant, aber nicht vorhanden.
3. Thema und Stichwörter: der Inhalt des Feldes sollte laut Empfehlung der Dublin-Core-Metadateninitiative aus kontrolliertem Vokabular entnommen werden, zum Beispiel allgemeingültigen Klassifikationen wie Dewey Decimal Classification oder anerkannten Fachthesauri.
In den Greenpeace-Datenbanken stehen die beiden Felder „Kategorie“ und „Stichworte“ zu diesem Zweck zur Verfügung, werden aber nicht mit Werten aus kontrolliertem Vokabular gefüllt. In den einzelnen Themen- und Bereichsdatenbanken ist zwar insofern eine gewisse Kontrolle über das jeweilige Vokabular vorhanden, als zum Zeitpunkt der Verschlagwortung keine neuen Stichwörter hinzugefügt werden können. Aber es liegt ganz in der Verantwortung der jeweiligen ThemenbereichsmitarbeiterInnen, wie sie die Stichwörter und Kategorien in ihrer Datenbank handhaben, sie können sie bei Bedarf erweitern oder verändern, es gibt keine zentrales kontrollierter Vokabular und damit auch keine Konsistenzprüfung. Hierin liegt ein Hauptgrund der inkonsistenten Erschließung.
Das Metadaten-Element ist also vorhanden, wird aber nicht in der empfohlenen Weise gefüllt. Dass es in zweifacher Ausprägung vorhanden ist,

⁷¹ „Auf der soziokulturellen Seite erfordern Ontologien .. die Einigung einer Gruppe von Anwendern auf die jeweiligen Begriffe und deren Zusammenhänge“, schreibt Staab in Bezug auf Ontologien; diese Aussage gilt meiner Ansicht nach für alle Formen kontrollierten Vokabulars, weil sie für eine Anwendergruppe dieselbe Funktion haben. (Staab 2002:200)

⁷² nach der deutschen Übersetzung des Dublin-Core-Elementesatzes, siehe <http://www.mpib-berlin.mpg.de/DOK/metatagd.htm> (Stand vom 12. 05. 2004)

⁷³ siehe dazu Anhang 2: Maske eines Dokuments in einer Bereichsdatenbank mit den Erschließungsfeldern

ist unschädlich, da ein Grundsatz des Dublin-Core-Standards heißt: Jedes Element ist wiederholbar.

4. Inhaltliche Beschreibung, „abstract“:
Dieses Feld ist bisher nicht vorhanden, es scheint aber auch kein Bedarf danach zu bestehen. Ein „abstract“ hat meist die Funktion, einen wissenschaftlichen Zeitschriftenartikel so kurz zusammenzufassen, dass ein Leser aus der Fachöffentlichkeit sich einen schnellen Überblick über die Relevanz des Artikels verschaffen kann. Bei den Dokumenten in den Greenpeace-Datenbanken handelt es sich dagegen um Dokumente, die meist in ihrer Gesamtheit wahrgenommen werden müssen, um einen sinnvollen Beitrag zum Wissensmanagement leisten zu können. In vielen Fällen sind die Dokumente auch von erheblich geringerem Umfang als Fachzeitschriftenartikel dies üblicherweise sind, so dass schon aus diesem Grund eine Zusammenfassung nicht notwendig wäre. Dieses Metadaten-Element wird daher als nicht relevant für die Greenpeace-Umgebung eingeschätzt.
5. Verleger oder Herausgeber: die Einrichtung, die verantwortet, dass diese Ressource in dieser Form zur Verfügung steht, wie z.B. ein Verlag, eine Universität etc.
Auch dieses Metadaten-Element ist in der Umgebung der internen Greenpeace-Datenbanken nicht relevant, die Dokumente sind (noch) nicht publiziert, sie stehen jedenfalls nicht zu Publikationszwecken in den Datenbanken. Falls sie publiziert werden, wäre für die große Mehrheit der Dokumente die Organisation Greenpeace e.V. Verleger oder Herausgeber, der Feldinhalt könnte also leicht automatisiert ergänzt werden.
6. weitere beteiligte Personen, die im Vergleich zum Verfasser sekundäre Beiträge geleistet haben.
Dieses Metadaten-Element scheint in der Greenpeace-Umgebung ebenfalls nicht relevant zu sein. Bisher existiert noch nicht einmal ein Feld für den primären Verfassernamen, es erscheint deshalb sehr fraglich, dass Bedarf nach diesem nachgeordneten Metadaten-Element besteht.
7. Datum, unterteilt zumindest nach Datum der Herstellung und Datum der letzten Änderung.
Beide Datumsfelder sind vorhanden, sie werden automatisch erzeugt.
8. Ressourcenart: sollte aus einer definierten Liste von Ressourcenarten⁷⁴ entnommen werden.
In den Greenpeace-Datenbanken ist ein Feld „Dokumentenart“ vorhanden, das die Funktion dieses Metadaten-Elements einnehmen kann. Wie bei Stichwörtern/Kategorien wird auch für die Inhalte in diesem Feld kein kontrolliertes Vokabular genutzt.
9. Format: sollte ebenfalls aus einem kontrollierten Vokabular⁷⁵ entnommen werden.

⁷⁴ <http://dublincore.org/documents/dcmi-type-vocabulary/> (Stand vom 12. 05. 2004)

⁷⁵ <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types> (Stand vom 12. 05. 2004)

Dieses Feld wäre sehr relevant für Zwecke der Langzeitarchivierung. Es ist in den Greenpeace-Datenbanken nicht vorhanden. Seine Inhalte könnten aber höchstwahrscheinlich mit vertretbarem technischen Aufwand automatisch erzeugt werden, da die möglichen Formate in der Intranet-Umgebung begrenzt sind und abgesehen von wenigen, zu vernachlässigenden Einzelfällen auch automatisiert auszulesen sein sollten. Da zudem der Themenbereich rund um die technischen Probleme der Langzeitarchivierung im Rahmen dieser Arbeit ausgeklammert bleibt, soll dieses Metadaten-Element im folgenden nicht als relevant bewertet werden.

10. Ressourcen-Identifikation: eindeutige Bezeichnung für die Ressource, z.B. URL, URI oder ISBN.

Eine eindeutige Identifikation von Dokumenten wird technisch gesehen im Rahmen der Lotus-Notes-Datenbankstruktur vorgenommen, es existieren so genannte Dokument-IDs, die ein Dokument im System eindeutig bezeichnen. Der Inhalt dieses Feldes wird auf der Oberfläche der Bereichs- und Themendatenbanken nicht angezeigt, das heißt er ist für die BenutzerInnen nicht transparent und auch irrelevant, da die Identifizierung von Dokumenten für BenutzerInnen nur in ihrer Funktion für z.B. Work-flow-Systeme oder Versionierungen interessant wäre. Ob danach Bedarf besteht, muss ermittelt werden.

11. Quelle: z.B. analoge Quelle digitalisierten Materials.

Nicht relevant, weil es sich bei den Dokumenten in den Greenpeace-Datenbanken um originär digitale Objekte handelt.

12. Sprache: die Sprache des Dokuments, sollte aus Liste von Sprachcodes⁷⁶ gewählt werden.

Ein entsprechendes Feld ist bisher nicht vorhanden, und es besteht zur Zeit auch kein Bedarf danach. Die Mehrheit der Dokumente liegt in deutscher Sprache vor, eine Minderheit ist englisch. Erst wenn die Datenbanken einem internationalen Benutzerkreis zugänglich gemacht würden – zum Beispiel anderen Greenpeace-Büros und Greenpeace International – wäre ein Bedarf nach einem Metadaten-Element für die Dokumentensprache vorhanden. In diesem Fall könnten die entsprechenden Feldinhalte wahrscheinlich mit vertretbarem Aufwand größtenteils automatisiert erzeugt werden. Deshalb wird dieses Metadaten-Element zur Zeit als nicht relevant einzuschätzen sein.

13. Beziehung zu anderen Ressourcen: Hinweise auf andere Dokumente, die einen Bezug zum beschriebenen Dokument haben.

Das Element ist nicht vorhanden, könnte aber relevant sein. Innerhalb des Lotus-Notes-Systems können in einem Dokument mit einfachen Mitteln Verknüpfungen erzeugt werden, die auf andere Dokumente, Dokumentensammlungen (so genannte Ansichten) oder ganze Datenbanken verweisen. Diese Möglichkeit wird von BenutzerInnen zunehmend genutzt, aber die Verknüpfungen werden in unstrukturierter Form irgendwo im Dokumentinhalt erstellt. Es könne also hier Bedarf nach Strukturierung durch ein zusätzliches Metadaten-Element bestehen.

⁷⁶ <http://www.loc.gov/standards/iso639-2/langhome.html> (Stand vom 12. 05. 2004)

14. räumliche und zeitliche Maßangaben: Auch hier wird empfohlen, Werte aus Listen kontrolliertem Vokabulars zu wählen. Dieses Metadaten-Element scheint in der Greenpeace-Umgebung nicht von Relevanz zu sein. Nur eine Minderheit der Dokumente hat Inhalte, die sich räumlich oder zeitlich beschreiben lassen, falls dies der Fall ist, sind die räumlichen Angaben Teil des thematischen Stichworts, zum Beispiel „Castor-Transporte Gorleben“ oder im Bereich der Koordination ehrenamtlicher MitarbeiterInnen „Gruppe Rhein-Main“. Eine Trennung von räumlichen Angaben und thematischem Stichwort ist in derartigen Anwendungsbereichen nicht sinnvoll, deshalb wird die Relevanz dieses Metadaten-Elements als gering eingeschätzt.

15. rechtliche Bedingungen: Urhebervermerk oder sonstige Aussage zu den Rechten, die an diesem Dokument bestehen, z.B. Lizenzbedingungen. In der Greenpeace-Intranet-Umgebung ist dieses Metadaten-Element nicht relevant, da es sich einerseits nicht um publizierte Informationen handelt, andererseits die Rechte an von Greenpeace-MitarbeiterInnen erstellten Dokumenten über arbeitsvertragliche Regelungen bei Greenpeace Deutschland e.V. liegen, also immer derselbe Feldinhalt einzutragen wäre. Für später mögliche gemeinsame Anwendungen z.B. mit anderen Greenpeace-Büros könnte der Feldinhalt also leicht automatisch generiert werden.

Zusammenfassend ist also festzustellen, dass folgende Metadaten-Elemente ergänzt oder ihre Handhabung verändert werden müssten, um eine mit dem Dublin-Core-Metadatenstandard konforme Erschließung in den Greenpeace-Datenbanken vornehmen zu können:

- Verfasserfeld zusätzlich
- Feldinhalte sollten aus Listen kontrollierten Vokabulars gewählt werden bei den Elementen
 - Thema/Stichwörter, also den Greenpeace-Datenbankfeldern Stichworte und Kategorien
 - Ressourcenart bzw. Dokumentenart

Ob noch weitere Elemente des Dublin-Core-Standards ergänzt werden sollten, weil danach von NutzerInnen-Seite Bedarf besteht, zum Beispiel nach einem Feld „Beziehung zu anderen Dokumenten“, muss durch eine Benutzerbefragung geklärt werden. Diese Befragung ist auch notwendig, um den gerade postulierten Bedarf nach einem zusätzlichen Feld und der Verwendung kontrollierten Vokabulars zu verifizieren.

3.3.3. Integration Metadaten/Suchmaschinentechnologie

Wie oben gezeigt, scheinen die zur Zeit zukunftssträchtesten Weiterentwicklungsansätze im Metadatenbereich darin zu bestehen, dass Suchmaschinentechnologie genutzt werden kann, um durch Metadaten erschlossene Informationsobjekte leichter auffindbar zu machen, als dies in den derzeit üblichen Volltextindizierungssystemen der Fall ist. Jay Jordan beschreibt zwei Projektzusammenhänge, in denen durch Dublin-Core-Metadaten erschlossene interne Dokumentensammlungen („repositories“) von Universitäten

in den USA und Großbritannien mithilfe von Suchmaschinentechnologie („harvesting“) effektiver nutzbar gemacht werden (Jordan 2004).

Eine derartige Strategie könnte auch für Greenpeace Deutschland Erfolg versprechend sein. Bevor sie im einzelnen diskutiert wird, ist zu prüfen, welche Voraussetzungen dafür vorliegen müssten. Folgende Komponenten müssten gegeben sein:

Die Greenpeace-Datenbanken müssten ein suchmaschinentaugliches Format haben, das heißt, die Dokumente müssten aus dem proprietären Lotus-Notes-Format in eine der Standard-Webauszeichnungssprachen HTML oder besser noch XML konvertiert werden können. Dies ist grundsätzlich möglich, da die Software Lotus Notes / Lotus Domino in den neueren Versionen eine immer weiter verbesserte Web-Integration bietet und entsprechende Funktionalitäten standardmäßig zur Verfügung stellt. Welchen Aufwand eine solche Migration verursachen würde, müsste allerdings noch im Detail geprüft werden.

Weiterhin müssten die Dokumente in konsistenter Weise durch Metadaten erschlossen sein. Dies könnte einerseits durch automatisierte Indizierung erreicht werden (siehe oben 3.2.1.) oder andererseits durch intellektuelle Erschließung mit den vorgeschlagenen Verbesserungen, die in der Anwendung des Dublin-Core-Standards zum Tragen kämen. Auch eine Kombination beider Verfahren wäre denkbar. Zunächst könnte als Basis ein Metadatenschema und kontrollierte Vokabularien wie z.B. ein organisationsweiter Thesaurus eingeführt werden, dann könnte versucht werden, eine retrospektive Verschlagwortung der existierenden Dokumente mithilfe von Verfahren automatischer Indizierung vorzunehmen. Dies wäre dann die Grundlage, auf der Suchmaschinentechnologie eingesetzt werden könnte.

Es wird deutlich, dass dies ein ehrgeiziges Projekt wäre, das vielfältiger Vorarbeiten bedarf. Andererseits könnte die Realisierung durchaus weniger aufwändig sein, als dies auf den ersten Blick erscheinen mag. Denn im Rahmen der oben erwähnten Projekte US-amerikanischer Universitäten ist schon viel Grundlagenarbeit geleistet worden, was zur Entwicklung von open-source-Software führte, die jedenfalls von ihrem Anspruch her universell einsetzbar sein soll. So ist als Basissoftware für institutionelle Dokumentensammlungen, die den Dublin-Core-Metadatenstandard benutzen, „Dspace“⁷⁷ in Zusammenarbeit von MIT und Hewlett Packard entwickelt worden. Darauf aufbauend könnten von OCLC im Rahmen der Open Archives Initiative entwickelten open-source-Produkte wie „OAIcat“ und „OAIHarvester“ eingesetzt werden⁷⁸. Andererseits darf nicht unterschätzt werden, dass der Funktionsumfang solcher Software zunächst genau untersucht werden muss und dann zu prüfen ist, ob und wie er mit den im konkreten Projekt festzustellenden Bedürfnissen der NutzerInnen in Einklang zu bringen ist.

⁷⁷ <http://www.dspace.org/> (Stand vom 12. 05. 2004)

⁷⁸ siehe Jordan 2004: „... Dspace, the open-source repository software created by MIT and Hewlett Packard. ...We have two open-source software programs – OAIcat and OAIHarvester – that implement OAI protocols for data storage and harvesting in support of institutional repositories.“

Zunächst müsste aber im Rahmen dieser Arbeit schon festgestellt werden können, ob die Voraussetzungen – Anwendung des Dublin-Core-Metadatenschemas, kontrollierte Vokabularien – für den Einsatz von Suchmaschinentechnologie im Greenpeace-Dokumentenmanagementsystem mit vertretbarem Aufwand geschaffen werden können und ob auf NutzerInnenseite Bedarf danach besteht.

3.4. Überprüfung der Ergebnisse durch Bedarfsanalyse: Benutzerbefragung

Schon 1972 wurde in einem informationswissenschaftlichen Grundlagenwerk festgestellt: „Jedes Informationssystem ist *jemandes* Informationssystem.“ (Kunz/Rittel 1972:29, Hervorhebung vom Verfasser) Grundsätzlich kann jedes Dokumentensuchsystem nur dazu dienen, von Menschen produziertes und in Dokumenten festgehaltenes Wissen anderen Menschen zugänglich zu machen (Ferber 2002:916). Auch in Bezug auf Metadaten wird in der Literatur die Relevanz des jeweiligen Benutzungskontexts betont:

„... metadata itself ... is not very helpful and doesn't get us very far. Metadata becomes interesting and useful when we employ it to construct information retrieval and management systems. ... The context of use is everything.“

(Lynch 2002:183)

Das bedeutet, dass die Ermittlung des konkreten Bedarfs, der bei den NutzerInnen vorliegt, vordringliche Aufgabe in jedem Projekt ist, das sich mit der Optimierung von Dokumentenmanagementsystemen befasst. Im Rahmen dieser Arbeit war unter den gegebenen zeitlichen Einschränkungen – und unter den Bedingungen eingeschränktem Zugriffs auf die Ressourcen anderer Greenpeace-MitarbeiterInnen – eine vollständige Bedarfserhebung nicht zu leisten. Ich habe mich deshalb auf eine Befragung repräsentativer NutzerInnen beschränkt, die ich in Form von Interviews durchgeführt habe; dies erschien im Vergleich zum Fragebogen, den die NutzerInnen selbständig ausfüllen, als die angemessenere Methode, denn es bietet einerseits den Vorteil höherer Verbindlichkeit, andererseits waren zur Bedarfsklärung offene, qualitative Fragestellungen notwendig, die im persönlichen Gespräch besser zu beantworten sind als in der asynchronen Kommunikation über einen Fragebogen.

3.4.1. Inhalte der Befragung

Die Befragung bezieht sich auf folgende Themen, die sich im bisherigen Verlauf der Diskussion als klärungsbedürftig erwiesen haben:

- Organisationsweites Klassifikationsschema, Fachthesaurus o.ä. – dies wäre sowohl für eine automatisierte Indizierung notwendig und könnte auch als kontrolliertes Vokabular für thematische Stichwörter / Kategorien und Dokumentenarten eingesetzt werden
- Verfasserfeld
- Verlinkung innerhalb des Datenbanksystems (Metadaten-Element „Beziehung zu anderen Ressourcen“)

- Bedarf nach optimiertem Retrieval nach Volltextsuche, sei sie nun auf Suchmaschinentechologie oder auf intern verbessertem Volltextindex beruhend
- Dokumentenidentifikation (Versionskontrolle, Work-flow)

Außerdem sollte mit der Befragung verifiziert werden, ob die im Projektstartdokument formulierten Problembereiche (siehe oben 3.1.) auch von den NutzerInnen mehrheitlich so wahrgenommen werden, ob also aktuell Bedarf nach Optimierung des Dokumentenmanagements in dieser Form festgestellt werden kann.

Aus diesen Themenfeldern wurde ein strukturierter Fragebogen⁷⁹ entwickelt, damit die Ergebnisse systematisch ausgewertet werden können. Selbstverständlich mussten die Fragen auf den Empfängerhorizont der BenutzerInnen zugeschnitten sein, es wurde beispielsweise nicht explizit gefragt, ob Bedarf nach einem organisationsweiten Klassifikationsschema besteht. Stattdessen wurden Fragen gestellt, die den Umgang mit den Erschließungsfeldern in der jeweiligen Bereichs- und Themendatenbank zum Gegenstand hatten. Daraus konnten Rückschlüsse auf die Konsistenz der Erschließung gezogen werden. Weiterhin wurde nach der Zufriedenheit der NutzerInnen mit der jetzigen Erschließungssituation bzw. nach einem möglichen Bedarf darüber hinaus gefragt.

Außerdem wurden Fragen nach der Benutzungshäufigkeit und -intensität gestellt. Dies ist notwendig, um eine grobe Einschätzung der Validität von Aussagen von NutzerInnen im Rahmen der Befragung vornehmen zu können. Denn ein Nutzer, der das Dokumentenmanagementsystem sehr selten und nur wenig intensiv nutzt, wird naturgemäß weniger Einblick in seine vielfältigen Funktionalitäten haben als jemand, der es intensiv zu seiner täglichen Tätigkeit braucht. Entsprechend steht zu vermuten, dass auch Bedarfsäußerungen unterschiedlich gewertet werden könnten.

Für die Details der einzelnen Fragen verweise ich auf die im Anhang beigefügten Fragebögen, die exemplarisch für zwei Befragungen stehen, in denen NutzerInnen sehr unterschiedliche Einschätzungen über die Erschließungssituation und entsprechend divergierenden Bedarf nach Optimierung genannt haben.

3.4.2. Zielgruppe / NutzerInnen

Wie oben in der Einleitung schon kurz erwähnt, ist die Greenpeace-Organisationskultur in außergewöhnlichem Maße auf flache Hierarchien und eigenverantwortliches Arbeiten der MitarbeiterInnen ausgerichtet. Zur Erreichung der Organisationsziele werden sehr unterschiedliche Fähigkeiten und Kompetenzen von MitarbeiterInnen und auch sehr unterschiedliche

⁷⁹ siehe Anhänge 5 und 6, ausgefüllte Fragebögen von zwei exemplarischen Befragungen

Kooperationsformen benötigt.⁸⁰ Daraus folgt ein weitgefächertes Spektrum von NutzerInnen Typen und auch eine sehr unterschiedliche Anwendung des Dokumentenmanagementsystems in den verschiedenen Organisationseinheiten.

Um eine für die Zwecke der Bedarfsanalyse repräsentative Stichprobe der Zielgruppe zu gewinnen, mussten also Kriterien aufgestellt werden, die für die Auswahl der Befragten tauglich waren, indem sie die Heterogenität operationalisierbar machen. Heterogenität der Zielgruppe ist im Bereich der Dokumentenmanagementsystem-NutzerInnen vor allem in diesen Ausprägungen vorhanden:⁸¹

- Hierarchieebenen und Tätigkeitsfelder der NutzerInnen (ein Geschäftsführer nutzt die Datenbanken aufgrund seiner Funktion wahrscheinlich auf andere Art als seine Assistenz, eine CampaignerIn nutzt sie anders als ein Mitarbeiter in der Telefonzentrale etc.)
- Typen von Organisationseinheiten, z.B. Themenbereich, Servicebereich⁸² (in einem Servicebereich, zum Beispiel der Personalabteilung, werden die Datenbanken wahrscheinlich anders genutzt als in einem so genannten Themenbereich, wo Kampagnen konzipiert und durchgeführt werden)
- unterschiedliche Datenbanknutzung, die sich beispielsweise anhand folgender Kriterien festmachen lassen könnte:
 - immanente Konsistenz der Erschließung der einzelnen Datenbank
 - Datenbank fungiert als operatives System mit Kooperationsfunktionen oder eher als Ablagesystem/Archiv
 - Benutzungshäufigkeit und -intensität des jeweiligen Nutzers.

Es darf nicht vernachlässigt werden, dass in Rahmen dieser Arbeit eine Vorklärung von möglichem Bedarf nach einem Realisierungsprojekt geleistet werden sollte. Deshalb schien es angemessen, auch so genannte „stakeholder“-Eigenschaften in der Stichprobe zu berücksichtigen, also auch darauf zu achten, dass RepräsentantInnen wichtiger Nutzergruppen und des Managements in die Befragung einbezogen werden. Teilweise ist dies schon durch das Kriterium der unterschiedlichen Hierarchieebenen gewährleistet. Es bedeutet aber darüber hinaus, dass mindestens eine/n VertreterIn der Geschäftsführung und mehrere Mitglieder der Bereichsleiter-Ebene und aus anderen Gremien in die Befragung einbezogen werden mussten.

⁸⁰ siehe auf <http://www.greenpeace.org/deutschland/?page=/deutschland/greenpeace/jobs/personalpolitik-bei-greenpeace> (Stand vom 12. 05. 2004) die Aufzählung der unterschiedlichen Fähigkeiten und Kompetenzen, die Greenpeace-MitarbeiterInnen für ihre Tätigkeit benötigen: „Fantasievolle Konfrontation, gewaltloser Widerstand, strategische Planung, fundierte Recherche, bildreiche Ausdruckskraft, genaue Budgetkalkulation, soziale Kompetenz, seriöse, solide Verwaltung, visionäres Denken, Mut, Geduld und Eigenständigkeit ... so unterschiedliche Fähigkeiten sind bei Greenpeace-Mitarbeiter/innen gefragt.“

Personalpolitik bei Greenpeace wird deshalb unter anderem von folgenden Zielen geleitet:

- „...die Unterschiedlichkeit, die Vielfalt von Mitarbeiter/innen mit ihren Persönlichkeiten, Visionen, kreativen Fähigkeiten und Talenten einzusetzen und zu fördern
- die Grundlage dafür zu schaffen, dass Individualist/innen in einem Team mehr bieten als die Summe ihrer Einzelleistungen
- sowohl kooperative als auch nonkonformistische Verhaltensweisen zu fördern...“

⁸¹ siehe Anhang 7, Projektdokument „Interviewpartner“

⁸² siehe Anhang 1, Organigramm von Greenpeace Deutschland

Die Berücksichtigung aller dargestellten Kriterien führte zu einer Liste von 16 Greenpeace-MitarbeiterInnen⁸³, mit denen im März und April 2004 Befragungen durchgeführt wurden.

3.4.3. Ergebnisse der Befragung

Wie oben schon erwähnt, sind im Anhang⁸⁴ zwei Fragebögen beigelegt, die exemplarisch für zwei Befragungen stehen, in denen NutzerInnen sehr unterschiedliche Einschätzungen über die Erschließungssituation und entsprechend divergierenden Bedarf nach Optimierung genannt haben:

Eine Pressesprecherin artikuliert klaren Bedarf nach einem konsistenten, organisationsweiten Klassifikationssystem („GP-Thesaurus“), da sie durch die unterschiedliche und inkonsistente Erschließung der verschiedenen Themendatenbanken in ihrer bereichsübergreifenden Arbeit in diversen Themen und Kampagnen nicht ausreichend unterstützt wird. Für sie sind die Dokumentensammlungen in den Themendatenbanken nicht ausreichend erschlossen.⁸⁵

Auf der anderen Seite des Meinungsspektrums steht die Personalbereichsleiterin, die das aktuell realisierte Konzept der Bereichs- und Themendatenbanken für richtig hält, da die jeweils unterschiedliche Erschließungspraxis Ausdruck des spezifischen Arbeitsstils eines Teams oder einer Arbeitsgruppe ist. Sie hält ein organisationsweites Klassifikationssystem für nicht sinnvoll, nach ihrer Erfahrung würde dies der Organisationskultur bei Greenpeace widersprechen und eher Widerstand bei den AnwenderInnen hervorrufen, da die MitarbeiterInnen sich und ihren Team-Arbeitsstil in einem organisationsweiten Schema nicht wiederfinden würden.⁸⁶

Die detaillierten Ergebnisse der Befragungen sind in einer Tabelle zusammengefasst, die als Anhang 8 der vorliegenden Arbeit beigelegt ist. Daraus ergeben sich folgende Anforderungen, die von den NutzerInnen an die Erschließung der Datenbanken und an die Optimierung des Dokumentenmanagements im allgemeinen gestellt werden:

Der Bedarf nach einem Verfasserfeld wird überwiegend bejaht. Neun von 16 Befragten sagen klar, dass dieser Bedarf besteht, der Rest macht teilweise keine eindeutigen Aussage dazu, teilweise hält er ein Verfasserfeld zwar nicht für sehr notwendig, aber auch nicht für schädlich.

Bei der Auswertung der Fragen zum Bedarf nach Optimierung der Volltextindizierung, konsistenterer Erschließung oder gar einem organisationsweiten Klassifikationsschema ergeben sich folgende Ergebnisse: Die große Mehrheit der befragten NutzerInnen ist zufrieden mit der Konsistenz der Erschließung in der eigenen Bereichsdatenbank (Spalte „Konsistenz“, Bewertung meist „hoch“, oft „mittel“). Hier wurde also eindeutig die oben aufgestellte

⁸³ siehe Anhang 7, Projektdokument „Interviewpartner“

⁸⁴ Anhänge 5 und 6

⁸⁵ Anhang 5

⁸⁶ Anhang 6

Arbeitshypothese, dass die mangelnde Konsistenz der Erschließungsbegriffe ein Problem in der Benutzung bewirkt, nicht bestätigt. Es scheint sich in den einzelnen Arbeitsgruppen jeweils ein dem spezifischen Bedarf angemessener Umgang mit dem unkontrollierten Vokabular der bestehenden Erschließungsfelder herausgebildet zu haben.

Beim möglichen Bedarf nach einem organisationsweiten Klassifikationsschema gehen die Meinungen sehr auseinander, sie bewegen sich zwischen den oben dargestellten Extremen der Pressesprecherin und der Personalchefin. Auch wenn die theoretische Möglichkeit einer besseren Datenbanknutzung durch konsistentere Erschließung anerkannt wird, wird oft die Frage nach dem Verhältnis von Aufwand und Nutzen eines solchen Projekts aufgeworfen. Höchstens in Teilbereichen, zum Beispiel nur innerhalb der Themendatenbanken im engeren Sinne (den Datenbanken der so genannten Themen- und Kampagnenbereiche, sozusagen dem „Kerngeschäft“ von Greenpeace⁸⁷) wird ein gemeinsames Klassifikationssystem für möglich gehalten. In Form einer „Metastruktur“ über die einzelnen Bereichs- und Themendatenbanken hinweg wird ein organisationsweites Klassifikationssystem von einer Minderheit der Befragten für möglicherweise nützlich gehalten, allerdings nicht als dringender aktueller Bedarf artikuliert.

Von mehreren NutzerInnen wird darauf aufmerksam gemacht, dass die persönliche Kommunikation über die Bereichsgrenzen hinweg wichtiger ist als optimierte Suchmöglichkeiten in den Datenbanken: Wenn ein Dokument in einem fremden Fachgebiet gesucht wird, würden auch eine konsistentere Erschließung oder optimierte Retrievalmöglichkeiten weniger nützlich sein als die jetzt schon in diesen Fällen praktizierte persönliche Information bei den Fachleuten im jeweiligen Bereich. Denn meist ist man nicht vertraut mit der Fachterminologie, so dass Volltextsuche wenig erfolgversprechend ist, und ein organisationsweites Klassifikationssystem würde wahrscheinlich in einiger Detailtiefe für fachfremde Personen auch nicht mehr aussagekräftig sein.

Hier zeigt sich, dass Methoden und Verfahren, die im Bereich wissenschaftlicher Fachinformation üblich und nützlich sind, nicht ungeprüft auf den Bereich des Wissensmanagements eines Unternehmens oder einer Organisation übertragen werden können. Denn während allgemein im BID-Bereich eher für eine anonyme Vielzahl von NutzerInnen möglichst vollständige Fachinformation nutzbar gemacht werden soll, ist beim organisationsinternen Wissensmanagement immer ein spezifischer Nutzerkreis vorhanden, der Informationsobjekte für spezifische Zwecke benötigt, die auch die Angemessenheit von Erschließungsmethoden beeinflussen können. Bei Greenpeace Deutschland scheint ein überwiegender Teil der hier befragten NutzerInnen der Meinung zu sein, dass die persönliche Kommunikation zielführender ist bei der Suche nach valider Information im Dokumentenmanagementsystem als optimierte Suchmöglichkeiten durch bessere Erschließung. Führt man sich die Größe der Organisation⁸⁸ vor Augen, nimmt das

⁸⁷ siehe Anhang 1, Organigramm von Greenpeace Deutschland

⁸⁸ Greenpeace Deutschland hat 175 fest angestellte Mitarbeiter, siehe <http://www.greenpeace.org/deutschland/?page=/deutschland/greenpeace/ueberuns/index#mitarbeiter> (Stand vom 12. 05. 2004), es gibt ca. 250 NutzerInnen des Dokumentenmanagementsystems, die Differenz kommt durch temporär mitarbeitende Ehrenamtliche, PraktikantInnen etc. zustande.

nicht wunder: In einem Themengebiet arbeiten jeweils überschaubare Personengruppen, man kennt sich persönlich und kann über direkte Kommunikation schneller und mit mehr Vertrauen auf die Qualität der Ergebnisse nach Informationsobjekten suchen, als wenn man nur auf Suchfunktionen im digitalen System angewiesen wäre.

Zusammenfassend kann also festgestellt werden, dass eine optimierte Volltextindizierung, konsistentere Erschließung der einzelnen Datenbanken oder gar ein organisationsweites Klassifikationssystem nur von einer Minderheit der NutzerInnen als aktueller Bedarf artikuliert wurden.

Ein Bedarf nach einer eindeutigen Ressourcen-Identifikation und den damit zusammenhängenden Möglichkeiten der Versionierung und strukturierter Verlinkung (Metadatenelement „Beziehung zu anderen Ressourcen“) ließ sich nur in geringem Umfang feststellen. Von zwei NutzerInnen wurde ein Bedarf nach Work-flow-Anwendungen für spezielle Bereiche (Erarbeitung von Druckvorlagen, Gremienorganisation) geäußert, der thematisch hier einzuordnen wäre, aber wegen seiner spezifischen Ausprägung hier nicht weiter verfolgt werden soll.

Ein Wunsch nach Integration der später entstandenen Dokumentensammlungen

- Contentmanagementsystem für die eigene Website (www.greenpeace.de)
 - von Greenpeace International betriebenes Intranetsystem
- konnte im Rahmen der Benutzerbefragung nicht bestätigt werden, er wurde wenig oder gar nicht artikuliert.

Der Bedarf nach einer Archivierungsmöglichkeit, die über die jetzigen Funktionalitäten der Datenbanken hinausgeht, konnte dagegen bei einer Mehrheit der Befragten festgestellt werden. Wie diese Archivierung genau aussehen soll, wurde unterschiedlich beschrieben. Überwiegend wurde eine zentrale, das heißt organisationsweite Struktur des Archivs für angemessen gehalten, also eine Abkehr vom Prinzip der bereichsweisen Erschließung in diesem Bereich befürwortet. Hier wäre also möglicherweise ein Anwendungsgebiet für ein organisationsweites Klassifikationssystem.

Als zusätzliche Anforderungen, die zu einer Optimierung des Dokumentenmanagements führen könnten, wurden folgende Themen genannt:

- ein technisches Problem der Anhangsbearbeitung⁸⁹ soll vordringlich gelöst werden;
- die Bedienbarkeit der Datenbanken soll verbessert werden, von einigen NutzerInnen wurden Anforderungen beschrieben, die eine Portallösung nahelegen könnten.

⁸⁹ Standardmäßig können die Dokumente in Lotus-Notes-Datenbanken beliebige Anhänge (Attachments) enthalten. Die komfortable Bearbeitung von häufig benutzten Dateitypen (Winword, Excel) ist zwar durch eine erweiterte Funktionalität möglich, aber in der aktuellen Lösung instabil und nur für wenige Dateitypen realisiert. In einer neuen, zur Zeit noch nicht eingesetzten Version der Software wird eine Funktionalität zur Anhangsbearbeitung standardmäßig enthalten sein, die auf eine Lösung dieses Problems hoffen lässt.

Zusammenfassend ist also zu sagen, dass von den zu Projektbeginn angenommenen Problembereichen in der Benutzerbefragung nur die fehlende Archivierungsmöglichkeit als Bedarf verifiziert werden konnte. Nicht bestätigt wurde ein Bedarf nach Optimierung in folgenden Problembereichen:

- Ressourcen-Identität (Versionierung, strukturierte Verlinkung zwischen verschiedenen Datenbanken)
- konsistentere Erschließung in den einzelnen Bereichs- und Themendatenbanken

Nur wenige der theoretisch erarbeiteten Lösungsansätze wurden von den NutzerInnen als einem aktuellen Bedarf entsprechend bestätigt:

- Optimierung der Volltextindizierung: kein Bedarf, entweder Zufriedenheit mit jetzigem Zustand oder persönliche Kommunikation effektiver als Suche im System.
- Anwendung des Dublin-Core-Metadatenstandards: Verfasserfeld sollte ergänzt werden, aber kein Bedarf nach organisationsweit kontrolliertem Vokabular für thematische Stichwörter und Ressourcenarten, da damit die jetzige bereichsspezifische Erschließung nicht mehr möglich wäre. Eine „Metastruktur“ über bestehende Bereichs- und Themendatenbanken hinweg wird nur vereinzelt gewünscht.
- Integration von Metadatenerschließung und Suchmaschinentechologie: siehe oben, aktuell kein oder nur vereinzelt Bedarf festzustellen.

Zur Zeit scheint also auf Seiten der NutzerInnen wenig Optimierungsbedarf hinsichtlich der Erschließung im Dokumentenmanagementsystem zu bestehen. Das ist einerseits ein positives Ergebnis, weil es Zufriedenheit mit der aktuell realisierten Dokumentenerschließung signalisiert. Andererseits sollte dieses Ergebnis nicht darüber hinweg täuschen, dass Benutzerbefragungen immer nur eine Momentaufnahme aktuell bestehenden, bewussten Bedarfs ergeben können, dass sie zum Beispiel keine Aussagen über die Zukunftssicherheit eines Systems machen können. Denn von NutzerInnen kann selbstverständlich nicht erwartet werden, dass sie das Optimierungspotential einer existierenden Lösung sachlich fundiert einschätzen können.

3.5. Empfehlungen / Lösungsvorschläge

Grundsätzlich muss berücksichtigt werden, dass das Dokumentenmanagementsystem bei Greenpeace Deutschland inhaltlich als Element eines Wissensmanagementsystems zu werten ist, also in einem größeren Zusammenhang eingebettet ist. Oben (3.1.) wurde festgestellt, dass die Bedeutung eines professionellen Wissensmanagements umso höher einzuschätzen ist, je kritischer die Ressource Wissen für den Erfolg der jeweiligen Organisation ist (Willke 2001:86). In der konkreten Umsetzung bedeutet das, dass hier Methoden der professionellen Projektplanung zum Einsatz kommen müssen und das Gesamtprojekt Wissensmanagement nicht in unverbundenen Einzelprojekten wie „Optimierung des Dokumentenmanagementsystems“, „Entwicklung eines Mentoring-Systems“ etc. aufgespalten bleiben sollte, wie es derzeit der Fall ist. Denn auf diese Weise kann keine klare gemeinsame Zieldefinition stattfinden, die einzelnen Projekte oder Aktivitäten können nicht aufeinander abgestimmt und koordiniert werden, letztlich wird

höchstwahrscheinlich kein erfolgreiches Wissensmanagementsystem zum Nutzen der Organisation langfristig implementiert werden. In der Literatur wird diese Situation, an der viele derartige Projekte in wissensbasierten Organisationen scheitern, unter dem Stichwort „Aktionismus“ diskutiert, es werden stattdessen Phasenmodelle als Einführungsstrategie für professionelles Wissensmanagement vorgeschlagen (Haun 2002:323 ff.).

Die erste konkrete Empfehlung wäre also:

- Einbettung des Projekts „Optimierung des Dokumentenmanagementsystems“ in ein umfassendes Wissensmanagementprojekt;
- in diesem Rahmen Definition der Ziele, die mit der Optimierung des Dokumentenmanagementsystems erreicht werden sollen.

Die zu definierenden Ziele würden unter anderem auch dazu dienen, folgende bisher offene Fragen zu beantworten:

- Wie ist der Bezugszeitraum? Was soll kurz-, mittel-, langfristig erreicht werden?
- Wird die Integration mit dem Contentmanagementsystem und/oder mit weiteren, von Greenpeace International geplanten Systemen angestrebt?

Denn ohne Planungssicherheit in diesen Fragen, die ja nicht allein von aktuellen und artikulierbarem Bedarf auf Seiten der NutzerInnen abhängen können, sondern auch in der Entscheidungskompetenz des Managements liegen, lässt sich keine seriöse Projektplanung zur Optimierung des Dokumentenmanagementsystems machen.

Die Integration der beiden anderen Dokumentensammlungen (Contentmanagementsystem, Intranetsystem von Greenpeace International) wäre perspektivisch ebenfalls in die Projektplanung einzubeziehen. Von Seiten der NutzerInnen wurde aktuell zwar wenig oder kein Bedarf zur Integration dieser Systeme artikuliert. Aber die Erfahrung zeigt, dass je länger derartige Systeme parallel bestehen und je intensiver sie genutzt werden, desto wahrscheinlicher Bedarf nach ihrer Integration entsteht – und sei es nur, um unnötige Doppelarbeiten zu vermeiden oder um durch verringerte Heterogenität im IT-System die Effektivität von Systemwartungsaufwänden erhöhen zu können.

Zudem sind auch zur Zeit bei Greenpeace International Projekte zur Optimierung des dortigen Dokumentenmanagementsystems in verschiedenen Planungs- und Realisierungsphasen, unter anderem ein Projekt „Digital Asset Management System“ (DAMS), ein Teilprojekt davon befasst sich mit Dokumentenmanagement, dort wird geplant, zur Erschließung der Dokumente den Dublin-Core-Metadatenstandard einzusetzen. Allerdings ist zum jetzigen Zeitpunkt noch nicht absehbar, welche Metadaten-Elemente genutzt werden sollen, ob und welches kontrollierte Vokabular eingesetzt werden soll, wie die Heterogenität der Sprachen in diesem Bereich behandelt werden soll, ob also beispielsweise mehrsprachige Thesauri entwickelt werden sollen oder ein sprachenunabhängiges Klassifikationssystem. Alle diese Detailplanungen müssten weiter fortgeschritten sein, um eine Aussage darüber treffen zu können, welche Weiterentwicklungen im Bereich des Dokumentenmanagementsystems bei Greenpeace Deutschland angestrebt werden sollten, um perspektivisch eine Möglichkeit der Systemintegration zu schaffen. Wichtig wäre zum jetzigen Zeitpunkt lediglich, keine Schritte in die falsche Richtung zu gehen, also bei Weiterentwicklungen des

Greenpeace-Deutschland-Systems im Blick zu behalten, dass sie auch zukunftssicher für eine Integration mit dem DAMS-Projekt bei Greenpeace International sein sollten.

Als zweite konkrete Empfehlung wäre zu nennen:

- Einführung des Dublin-Core-Metadatenstandards zur Erschließung der Dokumente in den Bereichs- und Themendatenbanken.

Denn wie oben gezeigt, hat sich dieser Standard zur Erschließung digitaler Informationsobjekte inzwischen international allgemein durchgesetzt und kann deshalb am ehesten Zukunftssicherheit garantieren. Es ist „wichtig, eine möglichst große Übereinstimmung mit existierenden Metadatenstandards anzustreben, um eine nachhaltige Benutzung von organisationsinternen Metadaten und eine möglichst einfache Integration von organisationsexternen, ‚fremden‘ Metadaten zu erreichen“ (Staab 2002: 201 f.). Jeder Lösungsansatz sollte sich also daran orientieren, dass er der Einführung des Dublin-Core-Metadatenstandards möglichst reibungslos den Weg ebnet.

All dies vorausgeschickt, lassen sich als Ergebnis der vorliegenden Arbeit folgende konkreten Maßnahmen vorschlagen – einerseits für eine Minimallösung, die nur die notwendigsten und kurzfristig realisierbaren Bereiche zum Inhalt hat, auf der anderen Seite für eine Maximallösung, die alles berücksichtigt, was sich zur Optimierung des Dokumentenmanagementsystems als sinnvoll erweisen könnte:

3.5.1. Minimallösung

Eine Minimallösung für die festgestellten Problembereiche könnte aus folgenden drei Komponenten bestehen:

1. In den Bereichs- und Themendatenbanken sollte ein Verfasserfeld eingeführt werden. Dies ist unabdingbare Voraussetzung für eine spätere Einführung des Dublin-Core-Metadatenstandards und wird auch aktuell von den NutzerInnen als Bedarf artikuliert.
2. Eine „kleine“ Archivlösung sollte projektiert und kurzfristig realisiert werden, sie könnte als „Kampagnenarchiv“ beginnen und folgende Komponenten haben:
 - eine einheitliche Archivdatenbank für alle Themen- und Kampagnenbereiche
 - Ablage der Kampagnendokumente nach Abschluss der jeweiligen Kampagne gemäß vereinbarten Regeln
 - konsistente Erschließungsbegriffe mit kontrolliertem Vokabular für thematische Stichwörter

Hiermit könnten schon Vorarbeiten für eine auch in anderen Projekten nützliche Liste kontrollierten Vokabulars zur thematischen Verschlagwortung geleistet werden, eine erste Version eines organisationsweiten Fachthesaurus für Greenpeace Deutschland könnte entwickelt werden. Zudem besteht auch aktuell von Seiten der NutzerInnen ein Bedarf nach einer Archivierungslösung, der hiermit ansatzweise befriedigt werden könnte.

3. Die Kompatibilität aller Veränderungen oder Weiterentwicklungen des Dokumentenmanagementsystems bei Greenpeace Deutschland mit den Entwicklungen und Planungen im Zusammenhang mit dem DAMS-Projekt bei Greenpeace International ist sicherzustellen.

3.5.2. Maximallösung

Als Maximallösung kommen folgende Maßnahmen in Frage, die zwar nicht alle in notwendigem Sachzusammenhang stehen, aber keineswegs einzeln willkürlich herausgegriffen werden sollten. Stattdessen wäre hier ein strukturierter Projektplanungszusammenhang im Rahmen von Wissensmanagement dringend zu empfehlen, siehe oben.

Folgende Maßnahmen wären zu erwägen:

Ein „Greenpeace-Application-Profil“ sollte entwickelt werden, das auf Grundlage des Dublin-Core-Metadatenstandards den spezifischen Informations- und Erschließungsbedarf einer internationalen Umweltschutzorganisation abbildet. Dabei ist unverzichtbar, dass Sachkunde über die inhaltlichen Greenpeace-Themen beispielsweise von CampaignerInnen in ihrem jeweiligen Fachgebiet eingebracht wird; außerdem ist notwendig, dass sich die künftigen NutzerInnen in einem strukturierten Prozess auf die jeweiligen Begriffe einigen und einen Konsens über Relationen, Zusammenhänge und die Details ihrer künftigen Benutzung finden. Dies sollte möglichst als gemeinsames Projekt in Zusammenarbeit mit Greenpeace International und anderen nationalen Greenpeace-Büros geschehen. Auf dieser Grundlage könnte dann ein organisationsweites Klassifikationssystem, ein mehrsprachiger Fachthesaurus oder eine Ontologie erarbeitet werden.

Dies könnte als Basis dienen, um bei Greenpeace Deutschland eine mit dem entwickelten Standard konforme „Metastruktur“ auf einer Ebene oberhalb der Bereichs- und Themendatenbanken zu entwickeln. Diese könnte darüber hinaus auch die Grundlage für eine automatische Indizierung bzw. Klassifikation sowohl der existierenden Dokumente in den Bereichs- und Themendatenbanken als auch anderer, bisher nicht integrierter Datenquellen sein. In diesem Zusammenhang könnte eine umfassende Archivlösung projektiert werden, die letztlich die Funktion eines „Greenpeace-Gedächtnisses“ hat und sowohl digitale als auch analoge Informationsobjekte integriert.

Eine Migration der zur Zeit im proprietären Lotus-Notes-Format vorliegenden Dokumentensammlungen ins XML-Format wäre zu erwägen, wenn sie für Zwecke der automatischen Indexierung bzw. zum Einsatz von Suchmaschinentechnologie oder auch aus Gesichtspunkten der Interoperabilität mit anderen Dokumentensammlungen notwendig wird. Damit würde auch eine zukunftssichere Trennung von Inhalt und Form der Dokumente realisiert, und es gäbe darüber hinaus auch schnell zu realisierende Vorteile der Kompatibilität beim Dokumentenaustausch in der Kommunikation mit externen Partnern.

Auf der Ebene des Lotus-Notes-Systems käme die Einführung einer Portallösung in Frage, um die intuitive Bedienbarkeit und allgemeine Nutzerfreundlichkeit des

Datenbanksystems zu verbessern und damit den NutzerInnen den Zugang zu den Dokumentensammlungen zu erleichtern. Ein „Portal“⁹⁰ in einem Unternehmen oder einer Organisation ist gekennzeichnet durch:

- den einheitlichen Einstiegspunkt: MitarbeiterInnen erhalten einen zentralen Zugang zu allen im Organisationskontext für sie wichtigen Informationen;
- Simplizität: einfache, intuitive Bedienbarkeit;
- leistungsfähige Suchwerkzeuge, die ebenfalls möglichst intuitiv bedienbar sind und sowohl in Form einer thematischen Liste („browse“) als auch in Form einer Suchmaschine („search“) vorliegen;
- Aggregation großer Informationsmengen verschiedener Provenienz, die strukturiert und aufbereitet zur Verfügung stehen;
- Personalisierung: Die Software erlaubt die Eingabe von persönlichen, abteilungs- oder aufgabenbezogenen Präferenzen, was zu einer Differenzierung der Portaloberfläche und spezifischen Selektion bestimmter Informationen dient.

(siehe Rösch 2001:240 f.)

Auch unabhängig von der Einführung eines bereichsdatenbankübergreifenden Erschließungssystems („Metastruktur“) könnte damit der jetzt schon zu beobachtenden Zersplitterung der Datenbanklandschaft bei Greenpeace Deutschland entgegengewirkt werden, weil die NutzerInnen zumindest die Datenbanken, auf die sie aktuell Zugriff haben und nutzen könnten, besser im Blick hätten. Außerdem stellte eine Portallösung einen guten Ausgangspunkt zur Integration bisher unverbundener Systeme wie dem Contentmanagementsystem dar.

Wenn darüber hinaus im Rahmen einer Einführung des Dublin-Core-Standards auch eine Metastruktur zur Erschließung der Bereichsdatenbanken entwickelt würde, wäre dies ein idealer Erschließungszusammenhang für alle anderen Informationsobjekte innerhalb der Portallösung. Denn das zu entwickelnde Klassifikationssystem soll in jeder durch das Portal umfassten Dokumentensammlung dieselbe Funktion haben, nämlich die Themenbereiche und Fachgebiete, auf denen Greenpeace tätig ist, für die MitarbeiterInnen in sinnfälliger Weise zu strukturieren und damit die Suche nach validen Informationsobjekten zu erleichtern.

4. Schlussbemerkung

Die in der Einleitung zugespitzt formulierten Zweifel an der Legitimation bibliothekarisch-dokumentarischer Bemühungen um Metadatenstandards hat die vorliegende Arbeit, wie ich meine, zur Genüge widerlegt. Zwischen den aktuellen Entwicklungstendenzen im World Wide Web und der Herausbildung von Metadatenstandards im BID-Bereich besteht kein Widerspruchs- oder Konkurrenzverhältnis, sondern im Gegenteil ist der Dublin-Core-Standard gerade eine Referenzanwendung der weiterentwickelten Web-Auszeichnungssprachen und von Resource Description Framework, den technisch-konzeptionellen Grundlagen des „Semantic Web“. Zudem kann ein Metadatenstandard auch in

⁹⁰ „Dass die von Microsoft usurpierte Windowsmetapher in der Portalmetapher eine logische Ergänzung fand, erscheint retrospektiv selbstverständlich“, meint Rösch zu recht (2001:237)

anderen Umgebungen als dem öffentlichen Internet, nämlich im Intranetbereich von Unternehmen und Organisationen, eine solide Grundlage für Interoperabilität und Zukunftssicherheit schaffen, wie die Diskussion der Optimierungsmöglichkeiten für das Dokumentenmanagementsystem bei Greenpeace Deutschland gezeigt hat.

5. Literaturverzeichnis

- Baker, Thomas: „Languages for Dublin Core“. In: D-Lib Magazine 4 (1998), 12.
<http://www.dlib.org/dlib/december98/12baker.html> (Stand vom 12. 05. 2004)
- Baker, Thomas: „A Grammar of Dublin Core“. In: D-Lib Magazine 6 (2000), 10.
<http://www.dlib.org/dlib/october00/baker/10baker.html> (Stand vom 12. 05. 2004)
- Baker, Thomas; Dekkers, Makx: „Identifying Metadata Elements with URIs. The CORES Resolution“. In: D-Lib Magazine 9 (2003), 7/8.
<http://www.dlib.org/dlib/july03/baker/07baker.html> (Stand vom 12. 05. 2004)
- Baxmann-Krafft, Eva-Maria; Lechner, Edith: „Normung im Bereich Information und Dokumentation“. In: Marianne Buder et al. (Hrsg.), *Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die fachliche Informationsarbeit*. 4. Ausgabe. München u.a.: Saur, 1997, S. 903 - 926.
- Bearman, David; Miller, Eric; Rust, Godfrey; Trant, Jennifer; Weibel, Stuart: „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), 1.
<http://www.dlib.org/dlib/january99/bearman/01bearman.html> (Stand vom 12. 05. 2004)
- Berners-Lee, Tim: *Axioms of Web Architecture: Metadata*. 1997.
<http://www.w3.org/DesignIssues/Metadata.html> (Stand vom 12. 05. 2004)
- Berners-Lee, Tim: *Semantic Web Road map*. 1998.
<http://www.w3.org/DesignIssues/Semantic.html> (Stand vom 12. 05. 2004)
- Berners-Lee, Tim: *What the Semantic Web can represent*. 1998.
<http://www.w3.org/DesignIssues/RDFnot.html> (Stand vom 12. 05. 2004)
- Berners-Lee, Tim: *Why RDF model is different from the XML model*. 1998.
<http://www.w3.org/DesignIssues/RDF-XML.html> (Stand vom 12. 05. 2004)
- Berners-Lee, Tim; Hendler, James; Lassila, Ora: „The Semantic Web“. *Scientific American*, May 2001, zitiert nach <http://www.w3.org/2001/sw/> (Stand vom 12. 05. 2004)
- Bray, Tim; Paoli, Jean; Sperberg-McQueen, C.M. (Hrsg.): *Extensible Markup Language (XML) 1.0*. 2. Aufl. Deutsche Übersetzung. 2002. <http://edition-w3c.de/TR/2000/REC-xml-20001006/> (Stand vom 12. 05. 2004)

- Burkart, Margarete: „Thesaurus“. In: Marianne Buder et al. (Hrsg.), *Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die fachliche Informationsarbeit*. 4. Ausgabe. München u.a.: Saur, 1997, S. 160 - 179.
- Caplan, Priscilla: *Metadata fundamentals for all librarians*. Chicago: American Library Association, 2003
- Dekkers, Makx; Weibel, Stuart: „State of the Dublin Core Metadata Initiative, April 2003“. In: D-Lib Magazine 9 (2003), 4.
<http://www.dlib.org/dlib/april03/weibel/04weibel.html> (Stand vom 12. 05. 2004)
- DIN-SICT-Papier „Strategie für die Standardisierung der Informations- und Kommunikationstechnik (ICT)“. Berlin: DIN, 2002
- Dobratz, Susanne; Liegmann, Hans; Tappenbeck, Inka: „Langzeitarchivierung digitaler Dokumente“. In: ZfBB 48 (2001), 6, S. 326 - 332.
- Dörr, Marianne; Enderle, Wilfried; Hauffe, Heinz: „Elektronische Publikationen und Informationsdienstleistungen“. In: Rudolf Frankenberger et al. (Hrsg.): *Die moderne Bibliothek. Ein Kompendium der Bibliotheksverwaltung*. München: Saur 2004, S. 381 - 417.
- Dunsire, Gordon: „Integrating Dublin Core / RDF records with MARC21 via the OCLC Connexion service at the Centre for Digital Library Research“. In: ZfBB 50 (2003), 4, S. 188 -192.
- Duval, Erik; Hodgins, Wayne; Sutton, Stuart; Weibel, Stuart: „Metadata Principles and Practicalities“. In: D-Lib Magazine 8 (2002), 4.
<http://www.dlib.org/dlib/april02/weibel/04weibel.html>. (Stand vom 12. 05. 2004)
- Eversberg, Bernhard: *Eine seltene Sache. Erwartung und Ernüchterung bei der thematischen Katalogsuche*. 2003. <http://www.allegro-c.de/regeln/cosarara.htm> (Stand vom 12. 05. 2004)
- Ferber, Reginald: „Dokumentsuche und Dokumenterschließung“. In: Peter Rechenberg und Gustav Pomberger (Hrsg.), *Informatik-Handbuch*. 3. Auflage. München u.a.: Hanser, 2002, S. 913 - 934.
- Fischer, Thomas: „Metadata Registries: Aktueller Stand und zukünftige Entwicklungen“. In: ZfBB 50 (2003), 4, S. 199 - 204.
- Fühles-Ubach, Simone: *Analysen zur Unschärfe in Datenbank- und Retrievalsystemen unter besonderer Berücksichtigung der Redundanz*. Berlin: Universität, Dissertation 1997. <http://www.ib.hu-berlin.de/%7Ewumsta/ubach/index.htm> (Stand vom 12. 05. 2004)

- Garrison, William: „Retrieval Issues for the Colorado Digitization Project's Heritage Database“, In: D-Lib Magazine 7 (2001) 10.
<http://www.dlib.org/dlib/october01/garrison/10garrison.html> (Stand vom 12. 05. 2004)
- Gladney, Henry; Bennett, J.L.: What Do We Mean by Authentic? What's the Real McCoy? In: D-Lib Magazine 9 (2003) 7/8.
<http://www.dlib.org/dlib/july03/gladney/07gladney.html> (Stand vom 12. 05. 2004)
- Goldfarb, Charles F.: *The Roots of SGM - A Personal Recollection*. 1996.
<http://www.sgmlsource.com/history/roots.htm> (Stand vom 12. 05. 2004)
- Gorman, Michael: „Metadata: Hype and Glory“. In: Wayne Jones et al. (Hrsg.): *Cataloging the Web. Metadata, AACR and MARC21*. Lanham u.a.: American Library Association, 2002, S. 179 - 181.
- Gruber, Tom: What is an Ontology? 1993. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (Stand vom 12. 05. 2004)
- Haller, Klaus; Fabian, Claudia: „Bestandserschließung“. In: Rudolf Frankenberger et al. (Hrsg.): *Die moderne Bibliothek. Ein Kompendium der Bibliotheksverwaltung*. München: Saur 2004, S. 222 - 261.
- Hampel, Thorsten: *Virtuelle Wissensräume. Ein Einsatz für die kooperative Wissensorganisation*. Paderborn: Universität, Dissertation 2001 [Elektronische Ressource]. <http://webdoc.sub.gwdg.de/ebook/e/2003/uni-paderborn/hampel/disserta.pdf> (Stand vom 12. 05. 2004)
- Hauer, Manfred: „Wissensressourcen zutage fördern. Digitalisierung von Aufsätzen und anderen Texten mit maschineller Inhaltserschließung am Beispiel der Vorarlberger Landesbibliothek Bregenz“. In: BuB 55 (2003), 3, S. 192 - 196.
- Haun, Matthias: *Handbuch Wissensmanagement. Grundlagen und Umsetzung, Systeme und Praxisbeispiele*. Berlin, Heidelberg u.a.: Springer, 2002
- Hengel, Christel: „Normdaten und Metadaten. Die Idee eines Internationalen Authority File“. In: ZfBB 50 (2003), 4, S. 210 - 214.
- Hoffmann, Heinz-Werner; Diedrichs, Reiner: „Elektronische Datenverarbeitung“. In: Rudolf Frankenberger et al. (Hrsg.): *Die moderne Bibliothek. Ein Kompendium der Bibliotheksverwaltung*. München: Saur 2004, S. 418 - 455.

- Horridge, Matthew: CO-ODE - A Collaborative Open Ontology Development Environment. In: D-Lib Magazine 10 (2004), 1.
<http://www.dlib.org/dlib/january04/01inbrief.html#HORRIDGE> (Stand vom 12. 05. 2004)
- Jordan, Jay: „Global Networking of Information – OCLC’s Strategy for the Future“. Proc. 7th International Bielefeld Conference, Bielefeld, 3. - 5. Februar 2004.
<http://conference.ub.uni-bielefeld.de/proceedings/jordan.pdf> (Stand vom 12. 05. 2004)
- Kaiser, Max; Lieder, Hans-Jörg; Majcen, Kurt; Vallant, Heribert: „New Ways of Sharing and Using Authority Information. The LEAF Project“. In: D-Lib Magazine 9 (2003), 11.
<http://www.dlib.org/dlib/november03/lieder/11lieder.html> (Stand vom 12. 05. 2004)
- Kamke, Hans-Ulrich; Zimmermann, Kerstin: „Metadaten und Online-Learning“. In: Information - Wissenschaft und Praxis, 54 (2003), 6, S. 345 - 348.
- Knorz, Gerhard: „Indexieren, Klassieren, Extrahieren“. In: Marianne Buder et al. (Hrsg.), *Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die fachliche Informationsarbeit*. 4. Ausgabe. München u.a.: Saur, 1997, S. 120 - 140.
- Krause, Jürgen; Niggemann, Elisabeth; Schwänzl, Roland: „Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken“. In: ZfBB 50 (2003), 1, S. 19 - 28.
- Kuhles, Gabor; Wefers, Sabine: „Verwaltung, Erschließung und Archivierung multimedialer Dokumente in UrMEL“. In: ABI-Technik 21 (2001), 3, S. 219 - 231.
- Kunz, Werner; Rittel, Horst: *Die Informationswissenschaften: Ihre Ansätze, Probleme, Methoden und ihr Ausbau in der Bundesrepublik Deutschland*. München, Wien: Oldenbourg, 1972. http://sova.sulb.uni-saarland.de/sova/volltexte/2000/31/pdf/kunz_rittel.pdf (Stand vom 12. 05. 2004)
- Lagoze, Carl: „The Warwick Framework. A Container Architecture for Diverse Sets of Metadata“. In: D-Lib Magazine 2 (1996), 7/8.
<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html> (Stand vom 12. 05. 2004)
- Leskien, Hermann: „Ein Zeitalter für Bibliotheken. Vielfältig gewandelte Rahmenbedingungen erfordern eine tiefgreifende Neuorientierung“. In: ZfBB 44 (1997), 1, S. 1 - 19

- Lightle, Kimberly; Ridgway, Judith: „Generation of XML Records across Multiple Metadata Standards“. In: D-Lib Magazine 9 (2003), 9.
<http://www.dlib.org/dlib/september03/lightle/09lightle.html> (Stand vom 12. 05. 2004)
- Lyman, Peter; Varian, Hal: *How Much Information, 2003*.
<http://www.sims.berkeley.edu/research/projects/how-much-info-2003> (Stand vom 12. 05. 2004)
- Lynch, Clifford: „Future Developments in Metadata and Their Role in Access to Networked Information“. In: Wayne Jones et al. (Hrsg.): *Cataloging the Web. Metadata, AACR and MARC21*. Lanham u.a. : American Library Association, 2002, S. 183 – 187.
- McCallum, Sally: „Library of Congress Metadata Landscape“. In: ZfBB 50 (2003), 4, S. 182 - 187.
- Miller, Eric; Hillmann, Diane: „Libraries and the Future of the Semantic Web: RDF, XML and Alphabet Soup“. In: Wayne Jones et al. (Hrsg.): *Cataloging the Web: Metadata, AACR and MARC21*. Lanham u.a.: American Library Association, 2002, S. 57 - 64.
- Mühlhäuser, Max: „Multimedia“. In: Peter Rechenberg und Gustav Pomberger (Hrsg.), *Informatik-Handbuch*. 3. Auflage. München u.a.: Hanser, 2002, S. 855 - 876.
- Nagamori, Mitsuharu; Baker, Thomas; Sakaguchi, Tetsuo; Sugimoto, Shigeo; Tabata, Koichi: „A Multilingual Metadata Schema Registry Based on RDF Schema“. Proc. International Conference on Dublin Core and Metadata Applications, Tokyo, 22 - 26. Oktober 2001.
<http://www.nii.ac.jp/dc2001/proceedings/product/paper-31.pdf> (Stand vom 12. 05. 2004)
- Niederée, Claudia: „Metadaten als Bausteine des Semantic Web“. In: ZfBB 50 (2003), 4, S. 193 - 198.
- Nohr, Holger: *Grundlagen der automatischen Indexierung*. Berlin: Logos Verlag, 2003
- Oehlschläger, Susanne: „Metadaten - Alter Wein in neuen Schläuchen? Abschlussworkshop des Projektes META-LIB und 1. Metadaten-Workshop der Arbeitsstelle für Standardisierung Der Deutschen Bibliothek“. In: ZfBB 50 (2003), 4, S. 179 - 181.
- O'Neill, Edward; Lavoie, Brian; Bennett, Rick: „Trends in the Evolution of the Public Web 1998 – 2002“. In: D-Lib Magazine 9 (2003), 4.
<http://www.dlib.org/dlib/april03/lavoie/04lavoie.html> (Stand vom 12. 05. 2004)

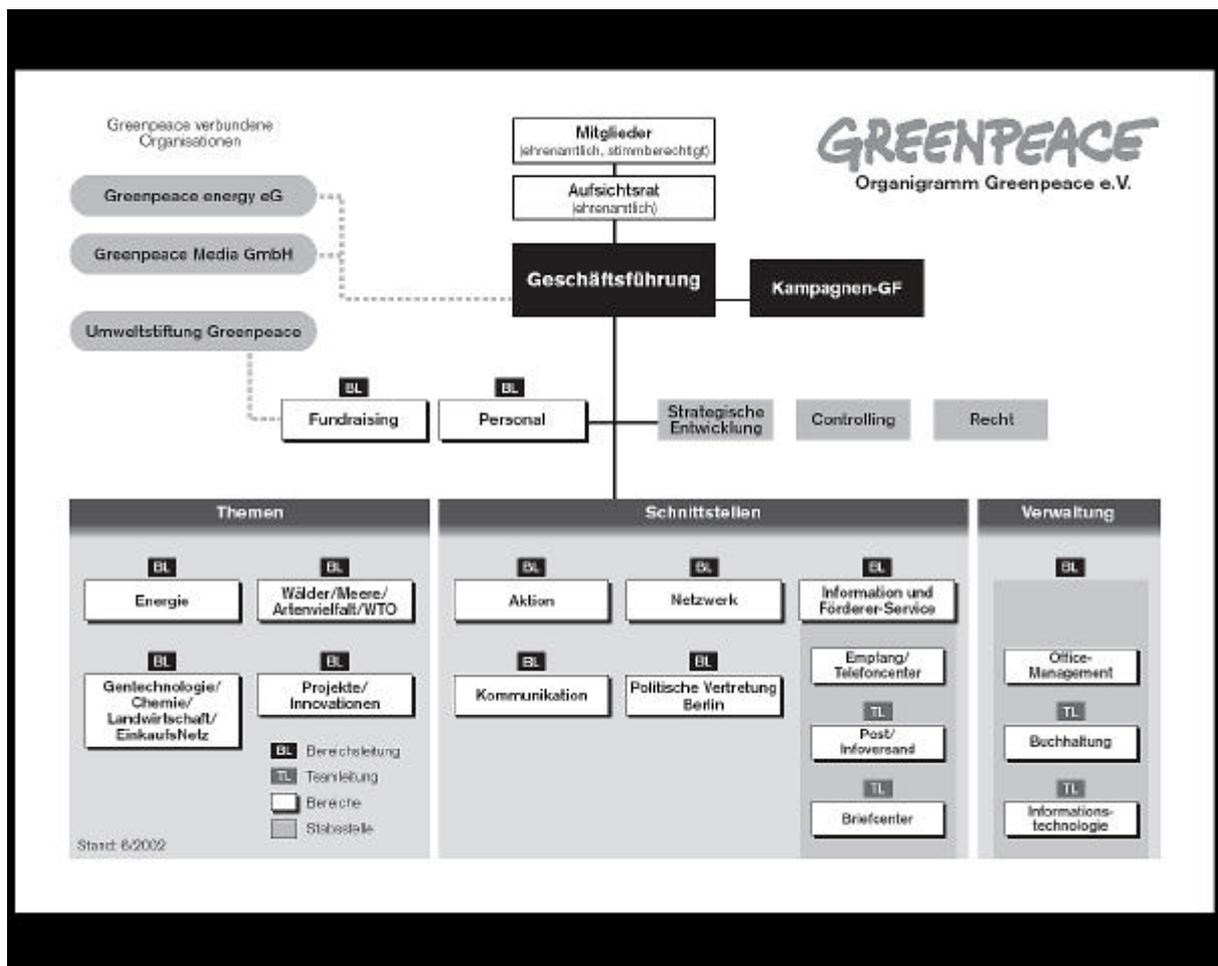
- Powell, Andy; Johnston, Pete: *Guidelines for implementing Dublin Core in XML*. 2003. <http://dublincore.org/documents/dc-xml-guidelines/> (Stand vom 12. 05. 2004)
- Rösch, Hermann: „Portale in Internet, Betrieb und Wissenschaft. Marktplatz und Instrument des Kommunikations- und Wissensmanagements“. In: B.I.T. online 4 (2001), 3, S. 237 - 246.
- Schiffer, Stefan; Templ, Josef: „Internetdienste“. In: Peter Rechenberg und Gustav Pomberger (Hrsg.), *Informatik-Handbuch*. 3. Auflage. München u.a.: Hanser, 2002, S. 1009 - 1024.
- Schwarz, Iris; Umstätter, Walter: „Die vernachlässigten Aspekte des Thesaurus: dokumentarische, pragmatische, semantische und syntaktische Einblicke“. In: nfd Information - Wissenschaft und Praxis 50 (1999), 4, S. 197 - 203.
- SGML Users' Group: *A Brief History of the Development of SGML*. 1990. <http://www.sgmlsource.com/history/sgmlhist.htm> (Stand vom 12. 05. 2004)
- Snijder, Ronald: *Metadata Standards and Information Analysis. A Survey of Current Metadata Standards and the Underlying Models*. 2001. <http://www.geocities.com/ronaldsnijder/> (Stand vom 12. 05. 2004)
- Staab, Steffen: „Wissensmanagement mit Ontologien und Metadaten“. In: Informatik-Spektrum 25 (2002), 3, S. 194 - 209
- Tappenbeck, Inka; Wessel, Carola: „CARMEN: Content Analysis, Retrieval and Metadata: Effective Net-working. Bericht über den middleOfTheRoad Workshop“. In: Bibliotheksdienst 35 (2001), 5, S. 566 - 572.
- Umstätter, Walter: „Wissenorganisation mit Hilfe des semiotischen Thesaurus - auf der Basis von SGML bzw. XML.“ In: Ohly, H.P.; Rahmstorf, G. und Sigel, A. (Hrsg.): *Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme*. Reihe: Fortschritte in der Wissensorganisation Bd. 6. Würzburg: Ergon, 2000, S. 237 - 249. <http://www.ib.hu-berlin.de/~wumsta/pub113.pdf> (Stand vom 12. 05. 2004)
- Weibel, Stuart: „Metadata: The Foundations of Resource Description“. In: D-Lib Magazine 1 (1995). <http://www.dlib.org/dlib/July95/07weibel.html> (Stand vom 12. 05. 2004)
- Wende, Ingo: „Normen und Spezifikationen der Informationstechnik“. In: Peter Rechenberg und Gustav Pomberger (Hrsg.), *Informatik-Handbuch*. 3. Auflage. München u.a.: Hanser, 2002, S. 1113 - 1132.
- Willke, Helmut: *Systemisches Wissensmanagement*. 2. Auflage. Stuttgart: Lucius und Lucius, 2001

Younger, Jennifer: „Metadata and Libraries: What's It All About?“ In: Wayne Jones et al. (Hrsg.): *Cataloging the Web: Metadata, AACR and MARC21*. Lanham u.a.: American Library Association, 2002, S. 3 - 11.

Zimmermann, Harald: „Möglichkeiten einer computergestützten Sacherschließung“. Proceedings 27. GfKI, Bibliothekarisches Programm, Cottbus, Juni 2003. http://archiv.tu-chemnitz.de/pub/2003/0066/data/zimmermann_saar.pdf (Stand vom 12. 05. 2004)

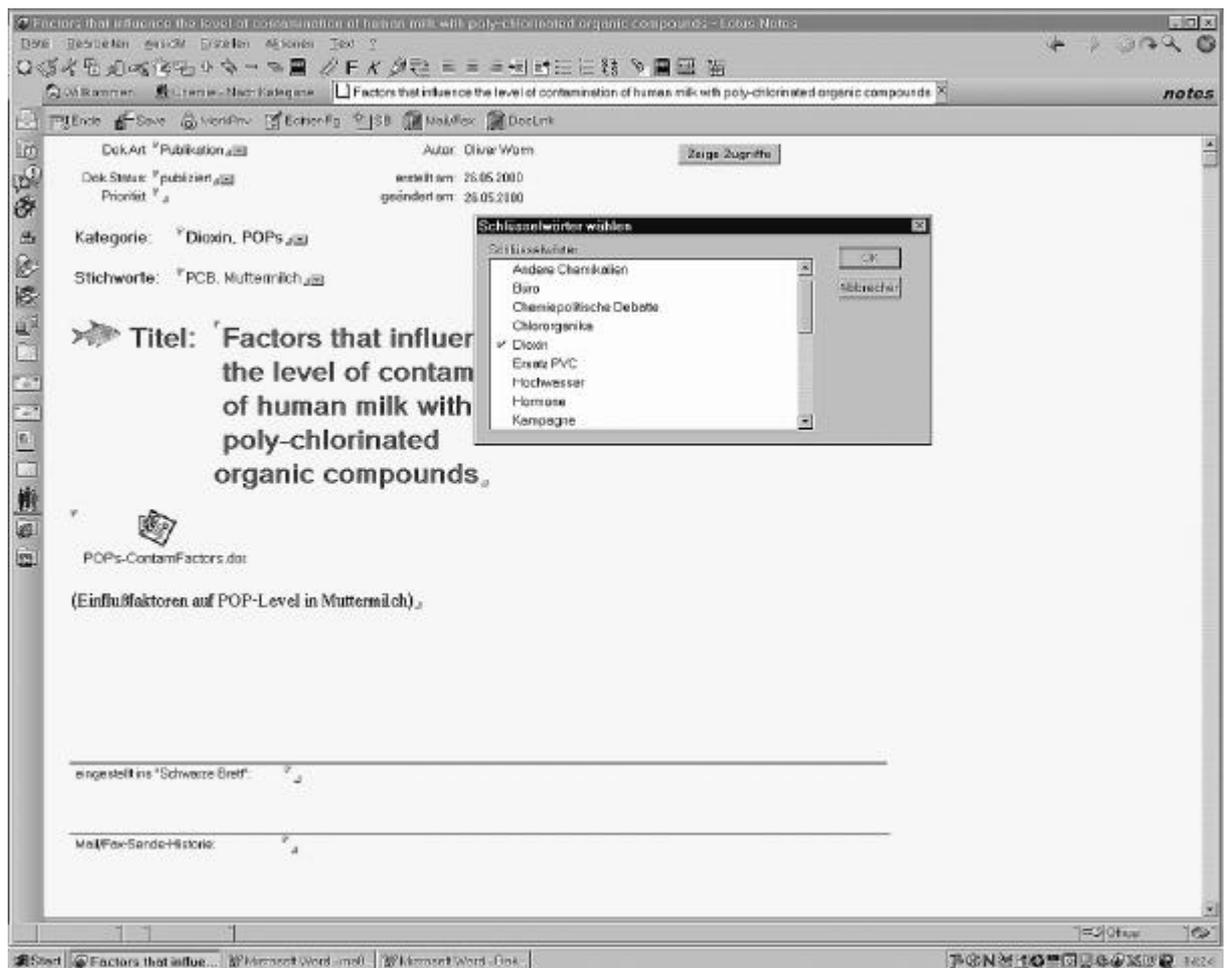
Anhang 1

Organigramm von Greenpeace Deutschland e.V.



Anhang 2

Maske eines Dokuments in einer Bereichsdatenbank mit Erschließungsfeldern:



Anhang 3

Suchfenster in einer Bereichsdatenbank

The screenshot shows a search window in a database application. The window title is 'Chemie - Nach Stichworten - Lotus Notes'. The search criteria are 'Suchen in: "Nach Stichworten"'. The search term is empty. The search options include 'Bedingung' (Condition), 'Autor' (Author), 'Nach Feld' (After Field), 'Maske' (Mask), 'Mehrere Wörter' (Multiple Words), and 'Beispielmaske ausfüllen' (Fill in example mask). The search options are set to 'Mit Wortvarianten (gut findet auch "gute")' (With word variants (good finds also "gute")), 'Ungenaue Suche' (Inexact search), and 'Ergebnisartierung: Nach Relevanz' (Result sorting: By Relevance). The search results are displayed in a table with columns for 'Titel' (Title), 'Autor' (Author), and dates. A dialog box 'Bedingung hinzufügen' (Add condition) is open over the results, showing 'Bedingung: Nach Feld' (Condition: After Field) and 'Dokumente suchen in denen das: enthält' (Search for documents in which it: contains) with the value 'PCB' entered.

Titel	Autor	Datum	Datum
Anschreiben an die Mitglieder des Umweltausschusses des Landes NRW	Manuel Fernandez		
PCB			
Muttermilchuntersuchungen der letzten 25 Jahre in Basel: Organochlorpestizid, PCB- und Nitromoschus-Rückstände	Oliver Worm		
Factors that influence the level of contamination of human milk with polychlorinated organic compounds	Oliver Worm		
Pesticides and Polychlorinated Biphenyl Residue in Human Breast Lipids and Their Relation to Breast Cancer	Oliver Worm		
Plasma Polychlorinated Biphenyl Levels in Dutch Preschool Children Either Breast-Fed or Formula-Fed during Infancy	Oliver Worm	26.05.2000	26.05.2000
Effects of environmental exposure to polychlorinated biphenyls and dioxins on cognitive abilities in Dutch children at 42 months of age	Oliver Worm	26.05.2000	26.05.2000
Umweltschadstoffe in Muttermilch	Oliver Worm	26.05.2000	26.05.2000
Effect of prenatal PCB exposure on Cognitive Processing Efficiency and Sustained Attention	Oliver Worm	05.01.1999	05.01.1999
Reduction in the Testosterone Levels by PCBs and DCE in Dall's Porpoises of Northwestern North Pacific	Oliver Worm	16.12.1998	16.12.1998
DDT and PCB Levels and Reproduction in Ringed Seal from the Baffin Bay	Oliver Worm	16.12.1998	16.12.1998
Environmental contaminants in tissues of a Neonate St Lawrence Beluga Whale	Oliver Worm	03.12.1998	03.12.1998
Concentrations of Polychlorinated Dibenzodioxins and Dibenzofurans and non-ortho substituted Biphenyls in Polar Bear Milk from Svalbard (Dioxin & PCBs in Milch von Eisbären auf Spitzbergen)	Oliver Worm	06.11.1998	06.11.1998
Chlorinated & brominated persistent organic compounds in historical samples from the	Oliver Worm	06.10.1998	06.10.1998

Anhang 4

Projektstartdokument

Dok.Art:
Dok.Status:
Priorität:

Autor: Gudula Holzheid
erstellt am: 25.02.2004
geändert am: 30.03.2004

Kategorie: Projekt\Studie Optimierung
Dokumentenmanagement

Ablageort:

Stichworte:

Titel: Projektstartdokument

Projekt Name : Studie zur Optimierung des
Dokumentenmanagements

Beginndatum : 01.03.04

Projekt Sponsor : G. M., Teamleitung IT

Projekt Team

Projektleiter : Gudula Holzheid

Technische Leitung : --

Fachliche Leitung : --

Andere : --

1. Überblick

1.1 Hintergrund Information

Seit 1996 setzt GPD Lotus Notes als Groupware ein, seit 1997/98 sind eigenentwickelte Lotus-Notes-Datenbanken im Einsatz, mit denen das interne Dokumentenmanagement betrieben wird.

Neben operativen Datenbanken, die z.B. Adressen enthalten, existieren zur Zeit folgende Datenbanken im Bereich des Dokumentenmanagements:

31 produktive Bereichs- und Themendatenbanken

(33 existieren und werden aktuell genutzt, 2 stehen zur Löschung an)

Diese Bereichs- und Themendatenbanken haben grundsätzlich die gleiche technische Gestaltung, die nur in Details voneinander abweicht. Daraus ergeben sich Vorteile hinsichtlich Wartungsfreundlichkeit und leichter Bedienbarkeit. Die Zugriffsberechtigungen sind bereichsspezifisch organisiert, d.h. ein Bereich oder eine Arbeitsgruppe hat Zugriff und vergibt bei Bedarf Zugriff an weitere Personen oder Gruppen.

diverse DBs mit speziellen Funktionen:

- MA-Info
- Meetings
- Veröffentlichungen
- GP-Handbuch / Schwarzes Brett
- dpa
- IT-Projekte

Diese Datenbanken haben ihre jeweils eigene, dem speziellen Zweck angemessene technische Gestaltung und stehen meist allen MitarbeiterInnen bei GPD im lesenden Zugriff zur Verfügung, der schreibende Zugriff ist je nach Datenbankfunktion unterschiedlich geregelt.

Die Datenbanken sind alle per Volltextindex erschlossen. Außerdem gibt es in jeder Datenbank bereichsspezifische Schlagworte, Kategorien etc., die zur Erschließung der Dokumente dienen. Die Kategorisierung bzw. Verschlagwortung der Dokumente wird zwar zum Zeitpunkt der Dokumentenbearbeitung mit kontrolliertem Vokabular durchgeführt, dieses Vokabular kann aber leicht und flexibel innerhalb jeder Datenbank erweitert werden und ist deshalb nicht wirklich als kontrolliertes Vokabular im dokumentarischen Sinne zu bezeichnen. Zudem existiert kein organisationsweiter Thesaurus oder ähnliches. Es handelt sich also um ein dezentrales, nicht hierarchisches Erschließungssystem, das sehr flexibel ist und den Bedürfnissen der einzelnen Themenbereiche oder Arbeitsgruppen entsprechend funktioniert.

Nach Jahren der Benutzung zeigen sich nun folgende Probleme:

- keine Archivierung
- keine bereichsübergreifenden Zugriffe
- dezentrale Lösung --> dasselbe Dok. in mehreren BTDBs vorhanden, evtl. in unterschiedlichen Versionen
- sehr unterschiedliche Nutzung in Bereichen --> Dokumentensammlungen sehr unterschiedlicher Konsistenz und Qualität sind entstanden
- keine Versionierung --> was ist aktuelles, gültiges Dok?
- trotz vorhandenem Dok.man.system Wissen immer noch weitgehend als implizites Wissen vorhanden, d.h. abhängig von Personen
- keine technische und/oder logische Verbindung zu:
 - Planet-CMS (www.greenpeace.de)
 - GL3-Intranet (www.nli.gl3) --> kein Zugriff für GPI u.a. NROs

1.2 Aussagen zum Projektumfang

1.2.1. Innerhalb des Projektumfangs

Studie, aufgrund derer über Realisierung der Empfehlungen entschieden werden kann.

1.2.2 Außerhalb des Projektumfangs

keine detaillierten Empfehlungen hinsichtl. technischer Lösungen
Realisierung der Empfehlungen.

1.3 Zeithorizont

Zeithorizont ist durch die Masterarbeitsfrist vorgegeben: 17. 05. 04 muss Projekt beendet sein.

2. **Ziele**

2.1 Angestrebtes Projektergebnis

Bestandaufnahme der Ausgangssituation

Dokumentenmanagement bei GPD

Beschreibung des Ist-Zustands

Vorteile

Nachteile

Dokumentenmanagement bei GPI

Beschreibung des Ist-Zustands

Bedarfsanalyse, d.h. Beantwortung folgender Fragen:

wie könnte Dokumentenmanagement bei GPD verbessert werden?

was wird von Anwenderseite bei GPD gewünscht?

was ist bei GPI geplant?

was ist aus informationswissenschaftlicher Sicht zu empfehlen?

Empfehlungen

Minimallösung

Maximallösung

Zusammengefasst soll das Projektergebnis also eine Entscheidungshilfe sein zur Frage: Wie weiter mit dem Dokumentenmanagement bei GPD?

2.2 Meilensteine

1. theoretische Vorarbeiten
(s. dazu Exposé und Gliederung Masterarbeit -->)
2. Bestandsaufnahme
 - 2.1. bei GPD
 - 2.2. bei GPI
3. Bedarfsanalyse
 - 3.1. Bedarfsermittlung bei GPD per stakeholder-Befragung
 - 3.1.1. stakeholders ermitteln
Termine mit stakeholders abstimmen
 - 3.1.2. Fragebogen konzipieren
 - 3.1.3. Fragebogen-Entwurf abstimmen mit GM
 - 3.1.4. Interviews führen
 - 3.1.5. Auswertung der Interviews
 - 3.2. Planungen bei GPI ermitteln
4. konkrete Empfehlungen ausarbeiten
(unter Einbeziehung theoretischer Vorarbeiten)
 - 4.1. minimale Lösung
 - 4.2. maximale Lösung

2.3. Qualitätssichernde Maßnahmen

- Auswahl Interviewpartner abstimmen mit GM
- Fragebogen-Abstimmung mit GM

2.4 Einführungsplan

entfällt

2.5 Datenschutz

Da die Studie im Rahmen einer Masterarbeit angefertigt wird, werden interne Dokumente außerhalb von GPD bekannt. Es ist zu prüfen, ob und inwieweit hierfür besondere Datenschutzmaßnahmen notwendig sind, ob z.B. Verschwiegenheitserklärungen eingeholt werden müssen.

3. Risiken

3.1 Annahmen zum Projekt

Keine ersichtlich

3.2 Liste möglicher Risiken

3.2.1 Technisch begründete Risiken

Keine ersichtlich

3.2.2 Durch das Projekt Team begründete Risiken

Ein-Personen-Projekt, zu wenig Input von Sach- und Fachkompetenz anderer Personen

Risikominierung

Qualitätssicherung durch Abstimmung mit GM.
Fachl. Leitung aus Anwenderkreis erst bei künftigem Realisierungsprojekt.

3.2.3 Andere Risiken

keine validen Ergebnisse der Befragung, weil nicht genügend Sachkunde bei InterviewpartnerInnen

Risikominierung

Fragebogen auf Empfängerhorizont zuschneiden

4. Ausstattung

keine Ausstattung über übliche Arbeitsmittel hinaus notwendig.

5. Budget

kein Budget für Sachmittel oder externe Dienstleistungen notwendig.

Anhang 5

Interview Personalleiterin

Autor: Gudula Holzheid
erstellt am: 15.03.2004
geändert am: 30.03.2004

Titel: Interview Personalleiterin

Datum: 24.03.04

Name: I. M.

Bereich: Personal

Funktion: BL

0. Datenschutz:

Hast du etwas dagegen, namentlich in meiner Masterarbeit erwähnt zu werden? Soll ich deinen Beitrag / Fragebogen anonymisieren?

ja

nein

1. Wie intensiv nutzt du:

1.1. Bereichsdatenbanken

täglich

mehrmals wöchentlich

mehrmals im Monat

nie

1.2. spezielle DBs

GP-Handbuch/Schwarzes Brett, MA-Info, Meetings-DB

täglich

mehrmals wöchentlich

mehrmals im Monat

nie

dpa, Veröffentlichungen

täglich

mehrmals wöchentlich

mehrmals im Monat

nie

2. wie suchst/findest du Dokumente in Notes? Über:

Volltextindex

immer

meistens

selten

nie

DB-Kategorien/Stichworte/Dokumentenarten...

immer

meistens

selten

nie

Suchreihenfolge:

1. Kategorien (--> dann in eigener DB meist gefunden)

2. Volltextindex

3. persönliche Kommunikation mit DB-Verantwortl.

3. Findest du die Kategorien/Stichworte in eurer DB...

gut, wie sie sind

verbesserungswürdig, weil...

unbrauchbar, weil...

kann ich nicht beurteilen

Verwaltungsbegriffe wurden vor ca. 1,5 Jahren überarbeitet.

4. Gibt es bei euch im Bereich Absprachen über die Verwendung von Kategorien/Stichworten...

überhaupt nicht, jede/r nach Lust und Laune

wir haben schon darüber gesprochen, setzen es aber nicht nicht sehr konsequent um

es gibt klare Absprachen, nämlich:...

weiß ich nicht

Die Verwaltungsbegriffe wurden in Teambesprechungen gemeinsam neu festgelegt, Zweifelsfälle wurden diskutiert und entschieden. Seitdem gibt es einen Konsens über ihre Verwendung im (kleinen) Personalbereich. Die Einarbeitung neuer MitarbeiterInnen ist unproblematisch, weil die DB-Struktur der Binnenstruktur des Bereichs entspricht, d.h. die Zuständigkeiten und Arbeitsgebiete darin abgebildet sind. Wenn neue Mitarbeiterinnen die Zuständigkeiten im Personalbereich kennengelernt haben, können sie damit auch die DB nutzen.

5. Wer erweitert bzw. verändert bei euch im Bereich die Kategorien/Stichworten? Nach welchen Prinzipien?

jede/r nach Lust und Laune

DB-Verantwortliche/r nach folgenden Prinzipien:

anderes Verfahren, nämlich:

weiß ich nicht

Wenn z.B. Arbeitsgebiet von mehreren Mitarbeiterinnen bearbeitet wird statt bisher nur von einer, wird es als DB-Verwaltungsbegriff aufgenommen, weil Dokumente zur gemeinsamen Bearbeitung abgelegt werden müssen.

6. Ist die fehlende Versionsverwaltung ein Problem für dich? Hast du z.B. schon mal mit einer veralteten Version eines Dokuments gearbeitet, ohne es zu bemerken?

ja

nein

Eigene Versionsverwaltung über Anhangsnamen, d.h. ein Notes-Dokument, mehrere Anhänge, aus deren Namen Bearbeitungsstand hervorgeht.

7. Gibt es in eurer DB Probleme mit der Authentizität von Dokumenten? Hast du dich z.B. schon einmal gefragt "Ist dieses Dok. tatsächlich von Person XY geschrieben, die als AutorIn erscheint?"

ja

nein

Verfasser-Feld wird vermisst.

8. Wir haben zur Zeit mehr als 30 Bereichs- und Arbeitsgruppendatenbanken bei GPD. Hast du deshalb schon mal nicht gewusst, in welcher DB du nach einem bestimmten Dokument suchen sollst?

ja

nein

Klare Abgrenzung der verschiedenen DBs, zumindest für I., die Initiatorin diverser Arbeitsgruppen-DBs ist.

oder umgekehrt: Stehst du manchmal vor dem Problem, dass dasselbe Dokument in mehreren Datenbanken vorhanden ist (Problem der Redundanz), möglicherweise in verschiedenen Versionen?

ja

nein

Aber Problem hat keine große Dimension, I. kann es selbst "im Kopf" behalten, wo sie Kopien von Doks abgelegt hat.

9. Wie findest du generell das Prinzip von Bereichsdatenbanken, auf die ein Bereich bzw. eine Arbeitsgruppe Zugriff hat?

gut, weil...

schlecht, weil...

keine Meinung

Gruppenzugehörigkeit z.B. bei Arbeitsgruppen/Steuergruppen kann gut darüber abgebildet werden, Zugriffe auf geheimhaltungsbedürftige Doks können differenziert vergeben werden.

Strukturierung über unterschiedliche Verwaltungsbegriffe je nach Thema und Umfang der Datenbank ist angemessen, unterschiedliche Erschließung ist Ausdruck des Arbeitsstils eines Teams bzw. einer Arbeitsgruppe

10. Es gibt keine logische oder technische Verbindung zwischen unserem internen Dokumentenmanagementsystem und dem Greenlink-Intranet (www.nli.gl3) dem Planet-CMS (www.greenpeace.de)

Vermisst du das in deiner täglichen Arbeit?

ja, in folgenden Situationen:

nein

GPI: Verbindung wäre nützlich, würde den Austausch mit GPI u.a. NROs vereinfachen. Keine Sprachbarriere für Personalbereich, weil wichtige Doks übersetzt. Zur Zeit schlechte Datenbasis von GPI, aber perspektivisch wäre gemeinsames System möglich und sehr sinnvoll.

Planet-CMS: kein Bedarf, die wenigen Dokumente (z.B. Stellenanzeigen), die vom Personalbereich auf www.greenpeace.de veröffentlicht werden, können auch auf den jetzigen Wegen weiterhin dorthin geschoben werden.

11. Anregungen, bisher nicht angesprochene Probleme, was könnte besser sein?

- Anhangsproblem lösen!
- Formatierungen aus Notes-Dokumenten sind nicht kompatibel mit anderen Textverarbeitungssystemen
- tendenziell plädiert Ingrid für eine Reduktion von Verwaltungsbegriffen (oder allgemein Möglichkeiten zur intellektuellen Indizierung) in Datenbanken
- ein Gesamtkonzept bzw. eine Standardisierung der Indexierung in BTDBs hält I. nicht für sinnvoll, nach ihrer Erfahrung würde dies der Organisationskultur bei GP widersprechen und eher Widerstand bei den AnwenderInnen hervorrufen, die MitarbeiterInnen würden sich und ihren Team-Arbeitsstil nicht in einer standardisierten DB-Struktur wiederfinden.

Anhang 6

Interview Pressesprecherin

Autor: Gudula Holzheid
Erstellt am: 15.03.2004
Geändert am: 30.03.2004

Titel: Interview Pressesprecherin

Datum: 05.03.04

Name: C. U.

Bereich: Kommunikation

Funktion: Pressesprecherin

0. Datenschutz:

Hast du etwas dagegen, namentlich in meiner Masterarbeit erwähnt zu werden? Soll ich deinen Beitrag / Fragebogen anonymisieren?

ja

nein

1. Wie intensiv nutzt du:

1.1. Bereichsdatenbanken

täglich

mehrmals wöchentlich

mehrmals im Monat, gelegentlich

nie

1.2. spezielle DBs

GP-Handbuch/Schwarzes Brett, MA-Info, Meetings-DB

täglich

mehrmals wöchentlich

mehrmals im Monat

nie

dpa, Veröffentlichungen

täglich

mehrmals wöchentlich

mehrmals im Monat

nie

2. wie suchst/findest du Dokumente in Notes? Über:

Volltextindex

immer in dpa-DB, Komms-DB

selten

nie
DB-Kategorien/Stichworte/Dokumentenarten...
immer
selten in Veröffentlichungen-DB,
nie

3. Findest du die Kategorien/Stichworte in eurer DB...

gut, wie sie sind
verbesserungswürdig, weil...
unbrauchbar, weil...
kann ich nicht beurteilen

In dpa- und Veröffentlichungen-DB sind GP-Themen als Kategorien
gutes Suchinstrument.

4. Gibt es bei euch im Bereich Absprachen über die Verwendung von
Kategorien/Stichworten...

überhaupt nicht, jede/r nach Lust und Laune
wir haben schon darüber gesprochen, setzen es aber nicht nicht sehr
konsequent um
es gibt klare Absprachen, nämlich:...
weiß ich nicht

Schreibende Zugriffe nur von H. T. (+ Praktikanten) --> sehr
konsistente Kategorisierung!

In anderen ("fremden") Bereichsdatenbanken, die C. im Rahmen von
Kampagnenarbeit nutzt, empfindet sie die Kategorisierung /
Verschlagwortung dagegen als stark verbesserungsbedürftig, wenn
nicht unbrauchbar. Schwierig ist für sie auch, dass jeder
Themenbereich andere Erschließungsprinzipien hat.

5. Wer erweitert bzw. verändert bei euch im Bereich die
Kategorien/Stichworten? Nach welchen Prinzipien?

jede/r nach Lust und Laune
DB-Verantwortliche/r nach folgenden Prinzipien:
anderes Verfahren, nämlich:
weiß ich nicht

s.o. Frage 4.: in Komms-DB okay, in dpa-DB sehr gut, in anderen
BTDBs sehr problematisch.

6. Ist die fehlende Versionsverwaltung ein Problem für dich? Hast du z.B.
schon mal mit einer veralteten Version eines Dokuments gearbeitet, ohne es
zu bemerken?

ja

nein

C. nutzt BTDB nicht als Arbeits-DB, sondern hat zusammen mit anderen Pressesprechern / Campaignern eigene "Versionsverwaltung" für Dokumentenaustausch per Mail abgesprochen.

7. Gibt es in eurer DB Probleme mit der Authentizität von Dokumenten? Hast du dich z.B. schon einmal gefragt "Ist dieses Dok. tatsächlich von Person XY geschrieben, die als AutorIn erscheint?"

ja

nein

Andererseits auch keine wichtige Information, dass Praktikant XY PE in Veröffentlichungen-DB gestellt hat, wichtiger wäre Name des/r PressesprecherIn.

8. Wir haben zur Zeit mehr als 30 Bereichs- und Arbeitsgruppendatenbanken bei GPD. Hast du deshalb schon mal nicht gewusst, in welcher DB du nach einem bestimmten Dokument suchen sollst?

ja

nein

C. vermisst eine DB-übergreifende Suchmöglichkeit.

oder umgekehrt: Stehst du manchmal vor dem Problem, dass dasselbe Dokument in mehreren Datenbanken vorhanden ist (Problem der Redundanz), möglicherweise in verschiedenen Versionen?

ja

nein

9. Wie findest du generell das Prinzip von Bereichsdatenbanken, auf die ein Bereich bzw. eine Arbeitsgruppe Zugriff hat?

gut, weil...

schlecht, weil...

keine Meinung

C. sitzt im Schnittstellenbereich Kommunikation als Pressesprecherin "zwischen allen Stühlen". Sie arbeitet in temporären Kampagnenteams, wo bereichsübergreifende Kampagnenproposal erarbeitet werden, in jedem Bereich gibt's unterschiedliche Arbeitsweisen und entsprechend ein unterschiedliches Ablagesystem, die Kategorisierung ist in jeder BTDB unterschiedlich, das bedeutet für C. eine aufwändige Einarbeitungsphase pro Kampagne.

10. Es gibt keine logische oder technische Verbindung zwischen unserem internen Dokumentenmanagementsystem und dem Greenlink-Intranet (www.nli.gl3) dem Planet-CMS (www.greenpeace.de)

Vermisst du das in deiner täglichen Arbeit?

ja, in folgenden Situationen:

nein

Greenlink: lesen reicht aus, keine weitere Verbindung nötig wegen Sprachbarriere

Planet: kein Bedarf

11. Anregungen, bisher nicht angesprochene Probleme, was könnte besser sein?

ein allgemeiner, bereichsübergreifende GP-Thesaurus wäre wünschenswert. Dieser müsste strukturiert sein nach Themen/Kampagnen und vor allem große Flexibilität bieten, da die Themen/Kampagnen sehr dynamisch sind.

Anhang 7

Liste der Interviewpartner

Dok.Art:
Dok.Status:
Priorität:

Autor: Gudula Holzheid
erstellt am: 25.02.2004
geändert am: 30.03.2004

**Kategorie: Projekt\Studie Optimierung
Dokumentenmanagement**

Ablageort:

Stichworte:

Titel: Interviewpartner

Kriterien für die Auswahl der InterviewpartnerInnen:

1. alle Hierarchieebenen und wichtigen Funktionen sollten vertreten sein:
GF
BL
Campaigner, Pressesprecher
Assistenz
2. alle Bereichstypen sollten vertreten sein:
Themenbereich
Schnittstellenbereich
Service-/Verwaltungsbereich mit Dokumentenmanagementbedarf,
z.B. Briefcenter, IT
3. diverse Typen von unterschiedlicher Datenbanknutzung:
unterschiedlich erschlossene DBs
(inkonsistente <-> konsistente Kategorisierung / Verschlagwortung)
unterschiedlich genutzte DBs
(operative DBs <-> Archivs-DBs)
unterschiedliches Maß der DB-Nutzung
(Power-User <-> gelegentliche DB-Nutzung)

Vorschlagsliste InterviewpartnerInnen:

Bereich	Name	Kriterium
GF	R. H.	1. GF 3. gelegentlicher DB-User
GF-Assistenz	A. S. E. S.-G.	1. GF-Assistenz 3. gelegentliche / häufige Anwenderin

Gentech/Chemie	S. F.	1. BL, Erfahrung als Campaigner 2. Themenbereich 3. Archiv-/operative DB (GentechWork/-Archiv)
Gentech/Chemie	M. K.	1. Campaigner 2. Themenbereich 3. Power-User
Gentech/Chemie	A. R.	1. Assistenz 2. Themenbereich 3. Archiv-/operative DB (GentechWork/-Archiv)
Meere/Wälder	J. S. T. L.	1. Assistenz, Referent 2. Themenbereich 3. wahrscheinlich konsistente, neu "aufgeräumte" BTDB
Netzwerk	D. K.	1. BL 2. Schnittstellenbereich 3. gelegentlicher DB-User
Netzwerk	C. G.	1. Campaignerin, Erfahrung als Assistenz 2. Schnittstellenbereich 3. ???
Briefcenter	K. L.-K.	1. TL 2. Verwaltungsbereich 3. operative DB
Personal	I. M.	1. BL 2. Verwaltungsbereich, Arbeitsgruppen 3. Power-User, konsistente Erschließung
Kommunikation	O. W.	2. Schnittstellenbereich 3. Power-User, eher inkonsistente DB?
Kommunikation	H. L.	2. Schnittstellenbereich 3. Power-User mit Dokumentationskenntnisse inkonsistente DB?
Kommunikation	C. U.	1. Pressesprecherin 2. Schnittstellenbereich 3. Power-User, bereichsübergreifend
Kommunikation	S. K.	1. Pressekoordinatorin, Erfahrung als Presses 2. Schnittstellenbereich 3. ???

Anhang 8: Tabelle Befragungsergebnisse

Name, Funktion	DB-Nutzung	Suche VT/EB	Konsistenz	Versionierung	Verfasserfeld	Redundanz	Einschätzung BTDB	sonstiger Beda Anm.
R. H. GF	selten, dpa-DB täglich	Kein Bedarf	k.A.	k.A.	nein	k.A.	+, Themen bed. DB-Struktur	Usability, Portal
E. S.-G. GF-Ass.	häufig	EB	hoch	nein	nein	nein	+, nicht noch mehr Information!	Anhangsproblem
A.S. GF-Ass.	täglich, sehr intensiv	Selten EB, oft VT	mittel	ja	k.A.	nein	+ pers. Komm.	Anhangsproblem, work-flow
I. M. BL Pers.	täglich, sehr intensiv	1. EB, 2. VT, 3.Komm	hoch	ja	ja	nein	+ Zugriffe, bereichsspez. Arbeitsweise	Anhangsproblem, Reduktion EB, 1 GPD-DB widerspricht Org.kultur
M. K. Camp.	mittel	VT	mittel	nein	nein	ja	+ Aufwand/ Nutzen?	Reduktion EB, Ablage zu aufwän
T. L. Referent	täglich, sehr intensiv	Meist VT, nur eigene Doks EB	hoch	ja	ja	ja	+ pers.Komm., Aufwand/ Nutzen?	Archiv
J. S. Ass.	täglich, sehr intensiv	Selten EB, meist VT	hoch	ja	ja	ja	- 1 DB für alle Themenber. machbar!	Archiv, Mail->DB zu aufwändig
D.K. BL NW	mittel	VT	hoch	k.A.	nein	nein	+ Zugriffe	Archiv
C. G. Camp.	täglich, sehr intensiv	Meist EB, Komm.	hoch	nein	nein	nein	+ Zugriffe, Arbeitsstil	pers. Komm. statt verschriftl. Information!
S. F. BL Gen	mittel	EB, VT ist „Notfunktion“	mittel	nein	ja	ja	+ Schnitt-mengen gering	Archiv, 1 GP-DB für publ. Doks, Mail->DB, Reduktion EB
A. R. Ass.	täglich	Meist EB	mittel	nein	ja	ja	+	Archiv
O. W. Recher- cheur	häufig	1. EB 2. VT	hoch	nein	ja	nein	+ DB-Struktur = Arbeits- struktur	Archiv, „Metasuche“
C. U. Presse- spr.	mittel, dpa-DB täglich	VT	mittel/ niedrig	nein	ja	ja	- „fremde“ DBs nicht benutzbar!	bereichsüberg. E wären sehr notwendig
S. K. Presse- spr.	täglich	1. EB, 2. VT Komm.	mittel/ hoch	ja	ja	nein	+ Zugriffe, Komm.	Archiv, work-flow, Rechtschreibpr., 1 GPD-DB nur mi Qualitätskontrolle
H. L. Video- Dokum.	mitel	Meist VT, nur eigene Doks EB	mittel	ja	ja	nein	+ 1 GPD-DB nicht möglich	Anhangsproblem, Bedienbarkeit
K. L-K. TL TC	täglich, sehr intensiv	Meist EB, zu viele VT-Erg.	mittel / hoch	ja	nein	nein	+ Arbeitsstil, Inhalte	Bedienbarkeit für Vielschreiber, Rechtschreibpr.