

Abschlussbericht zum Projekt: „Politisches Internet-Archiv“

1. Allgemeine Angaben

Name des Projekts:
Politisches Internet-Archiv

DFG-Aktenzeichen:
GZ:554922 (3) FES Bonn BIB44B0fes 03-01

Internetadresse des Vorhabens:
<http://www.fes.de/archiv/spiegelung/default.htm>

Mit dem Erstantrag kalkulierte Gesamtdauer des Vorhabens:
24 Monate

Projektbeginn / voraussichtlicher Abschluss der Arbeiten:
1. 09.2004 / 31.08 2006 (30.9.2006)

Erstbewilligung der DFG für dieses Projekt vom:
25.05.2004

Derzeit laufende DFG-Bewilligung vom:
25.05.2004

2. Angaben zum Projekt, Projektverlauf und zu den im Berichtszeitraum erreichten Projektergebnissen

2.1 Berichtszeitraum von: 1.09.2004 bis: 30.9.2006

2.2 Projektdarstellung

Ziel des Projekts:
Erfassung, Erschließung und Sicherung von Websites politischer Parteien der Bundesrepublik Deutschland sowie ihrer Fraktionen in den Parlamenten.

Die Websites politischer Großorganisationen (hier: politische Parteien) erlangen zunehmende Bedeutung für die Selbstdarstellung und für die politische Auseinandersetzung; dieses Medium entzieht sich – nicht zuletzt wegen der Integration multimedialer Bestandteile – bisher der archivischen Erfassung, Erschließung und Sicherung; diese müssen aber geleistet werden, um diese neue Quellengattung für die Parteienforschung, die schon jetzt das online-Angebot der politischen Parteien zum Gegenstand zahlreicher aktueller (Wahlkampf-)Analysen macht, auf Dauer zu erhalten. Als neuartige Quellengattung stellt das Internet zahlreiche neue Herausforderungen an die Archivierung auch der Internetpräsenzen der politischen Parteien. Schon der Einsatz herkömmlicher Informationstechnologien stellte und stellt die Archive vor nicht unerhebliche Schwierigkeiten, vor allem weil im Zuge der digitalen Generierung von Archivgut grundlegende Definitionen konventioneller Archivarbeit – wie z.B. die des Dokumentes oder der Provenienz – in Frage gestellt werden. Die Vielfalt der im Internet präsentierten Formate und Strukturen potenziert diese Problemlage erheblich. Es gilt nicht nur Text- und Bilddateien sowie die immer größer werdende Zahl von

Audio- und Videodateien zu erfassen, sondern diese im Zusammenhang der sie präsentierenden Strukturen auf Dauer zu sichern und der Forschung zur Verfügung zu stellen. Damit zukünftige Forscherinnen und Forscher diese Internetseiten aufrufen und sich in ihnen bewegen können, als wären sie direkt im Internet, müssen zahlreiche Probleme der Erfassung (Spiegelung), der Erschließung und der Präsentation gelöst werden. Erst die Lösung dieser Probleme unter den Aspekten der Authentizität, der Recherchierfähigkeit, Langfristigkeit und Benutzbarkeit eröffnet die Möglichkeit zum Aufbau eines Internet-Archivs.

Entwicklungsphase I (4 Monate)

In der ersten Phase geht es vorrangig um die Entwicklung von Verfahren zur Dynamisierung und Automatisierung des Spiegelungsprozesses. (AdsD)

Entwicklung von Kriterien für die Aufnahme bestimmter Seiten und Formate in die Spiegelung. Grundlage der Kriterienbildung soll sein, dass nur die Websites der Organisationsebenen und Personen gespiegelt werden, deren Schrift- und Sammlungsgut im Fokus der Sicherung und Betreuung des jeweiligen Archivs steht; über diesen Schwerpunkt hinausführende Links werden nicht verfolgt. (ADL)

Erprobung von kontinuierlichen Spiegelungen. Konsequenzen für die Präsentation und die Erschließung. (AdsD/ACDP)

Erprobung alternativer Erfassungsverfahren, - z.B. FTP (ADL/ACDP)

Feststellung der Grenzen von Spiegelungen. Überprüfung bisheriger Ergebnisse, z.B. Datenbanken, Javascripte, Flash (ACSP)

Deep Web: Ausloten von Möglichkeiten der Erfassung von besonders geschützten Webbereichen, z.B. Intranets, Passwortgeschützter Servicebereiche (AGG)

Entwicklungsphase II (11 Monate)

In einer zweiten Phase, in der es vor allem um Erschließung, Sicherung und Präsentation der archivierten Internetseiten geht, werden in arbeitsteiliger Vorgehensweise speziell folgende Problemfelder bearbeitet:

Indexierung: Ermöglichung von synchroner und diachroner Recherche über die gespiegelten Seiten. Überprüfung der Einbeziehung von ‚Wissensmanagement-Verfahren‘ (AdsD)

Klärung von Fragen der Langzeitarchivierung von Websites im Präsentationsformat: Bänder / Magneto-Optische Verfahren (ACSP)

Konversion in ein Speicherformat / Lösung der Migrationprobleme, die vor allem durch lange, konventionswidrige Dateinamen verursacht werden (AdsD / ADL)

Steigerung der Recherchemöglichkeiten durch die Entwicklung von modellhaften Verfahren zur Migration von speziellen Formaten (Pressedienste/Bilddokumente/Video- und Tondokumente) aus gespiegelten Seiten nach qualitativen Kriterien. (ACDP)

Präsentationsformen der Projekte: Server/CD (AdsD)

Überprüfung der Notwendigkeit bzw. Möglichkeit der Erfassung von Metadaten sowie Entwicklung von Verfahren zur Erfassung von Metatags und zur Erschließung über

Metatags. Diskussion des Dublin Core. Relevanz der einzelnen Elementen für die Erschließung von Seiten politischer Parteien (AGG)

Workshop zur Übertragung erarbeiteter Lösungen

Aufgaben des Workshops sind: Zusammentragen und Vorstellung der Ergebnisse. Diskussion der Übertragbarkeit einzelner Lösungen in unterschiedliche IT-Umgebungen. Interdependenzen zwischen unterschiedlichen Erfassungs-, Erschließungs- und Sicherungsmethoden. An diesem Workshop sollen zudem Vertreter(innen) von Archiven sowie Historiker(innen) und Politikwissenschaftler(innen) teilnehmen, die an einer Internetspiegelung im Hinblick auf die archivische Sicherung und archivarische Bearbeitung sowie auf die wissenschaftliche Nutzung dieser Quelle interessiert sind.

Kooperationsphase (9 Monate)

Entwicklung von Übertragungsmöglichkeiten von gefundenen Lösungen auf die jeweiligen technologischen Gegebenheiten. Untersuchung der Möglichkeit von Anbindungen an die jeweilige Datenbanken.

Gemeinsame Entwicklung von modellhaften Erschließungskriterien, Erfassungsmasken sowie Zitationsweisen. (Eventuelle Korrektur der technischen Verfahren)

Entwicklung der jeweiligen Standards für die Weiterführung des Projekts. Schulung der entsprechenden Mitarbeiter.

Die Publizierung der Ergebnisse, z.B. im „Archivar“ und auf Fachtagungen, z.B. beim Archivtag, kann in dieser Phase beginnen, wird aber in Eigenleistung der beteiligten Archive über die Projektdauer hinaus fortgeführt.

2.3 Arbeitsschritte im Berichtszeitraum

Projektdurchführung:

In allen beteiligten Archiven konnten neue Internet- Archive erfolgreich angelegt und ausgebaut werden. Zum festen Kern der Archivierung gehören die Websites der jeweiligen Parteien auf Bundes- und Landesebene sowie der entsprechenden Landtagsfraktionen. Dazu gehören auch die Sites wichtiger Funktionsträger in Partei und Fraktion. Einen Schwerpunkt der Arbeiten bildeten die Dokumentation der Wahlkämpfe auf Bundes- und Landesebene, die mit erheblichem Aufwand betrieben wurde.

Darüber hinaus haben einzelne Archive damit begonnen auch die Sites der

- Kreis- und Ortsvereinsebenen
- innerparteilichen Strömungen
- Vereinigungen und Sonderorganisationen
- der Vorfeldorganisationen sowie
- der Europaebene mit einzubeziehen

insgesamt wurde im Berichtszeitraum ein Datenvolumen von über 500 Gigabyte archiviert und gesichert. Je nach Zuschnitt der einzelnen Projekte wurden zwischen 100 und 500 Spiegelungen pro Archiv durchgeführt.

Da die Verzeichnung in der Datenbank Faust zum Projektstandard gehört, stehen die Internet-Archive prinzipiell den Archivbenutzern zur Verfügung, auch wenn sie zur Zeit überwiegend noch von den Archivkollegen und –kolleginnen genutzt werden. Ein Umstand, der sich mit der quellenmäßigen Abdeckung eines größeren Zeitraums sicher ändern wird.

Projektentwicklung:

- Erfassung

Standard

Die Spiegelung der Websites erfolgt in diskreten Schritten unter Verwendung eines Off Line Browsers. Dabei werden servergestützte Formate umgewandelt und die absoluten Links in relative umgeschrieben. Die gespiegelten Dateien werden zunächst in einem einheitlichen Verzeichnis auf einer Festplatte gespeichert.

Von allen Archiven wird der Off Line Explorer von Metaproduct genutzt. Die kommerzielle Software liefert immer noch die besten Ergebnisse, zeichnet sich durch einen großen Bedienungskomfort aus, der auch Eingriffe während des Spiegelungsprozesses zulässt. Er erlaubt die Verwendung von Macros und URL Substitutes, etwa um die Mehrfachspiegelungen von Dateien mit verschiedenen Session Ids zu vermeiden sowie die Verwendung kleinerer Scripte, um Spiegelungen vorzuprogrammieren und zu einem späteren Zeitpunkt oder in festen Intervallen automatisch durchführen zu lassen. Da keine proprietären Formate erzeugt werden, lassen sich die Spiegelungsergebnisse auch unterschiedlicher Off Line Browser miteinander kombinieren. Das gilt auch für die unter Umständen notwendige Ergänzung von Spiegelungen durch Streaming Files, die nur mit einer speziellen Software, z.B. dem RM-Recorder, erfasst werden können.

Erweiternde Methoden

In Auseinandersetzung mit den Ergebnissen des DFG-Projekts hat die Firma OIA ein Webarchivierungssystem entwickelt, das sich im wesentlichen durch die Einbeziehung einer relationalen Datenbank auszeichnet. Der von uns benutzte Off Line Browser ist in das System integriert. Die Datenbank soll eine redundanzfreie Archivierung der einschlägigen Websites garantieren. Dadurch würde aber auch ein quasi kontinuierlicher Spiegelungsprozess ermöglicht, der zwar noch in diskreten Schritten organisiert wäre, aber in beliebig dichten Intervallen erfolgen könnte. Auch die Analyse der im Archiv vorhandenen Dateien und Dateitypen, die jetzt noch einen eigenen Arbeitsschritt erfordert, würde vom System übernommen.

Das Webarchivierungssystem ist noch in der Entwicklung und konnte bisher nur eingeschränkt auf dem 9. Workshop des Projekts vorgestellt werden. Eine Teststellung im ACDP ist in Vorbereitung, eine im AdsD ist geplant.

Ergänzende Methoden

Um auch Bereiche des Deep Web zu archivieren, wurde vom Archiv GG mit der Übernahme der Daten aus dem verwendeten Content Management System (CMS) eine eigene Strategie entwickelt. Die gleichzeitige Übernahme der einschlägigen Metadaten bietet weitere Möglichkeiten des Zugriffs und der Recherche. Dabei wurde der im CMS verwendete Metadatensatz nach Dublin Core erweitert.

Ersetzende Methoden

Andere Möglichkeiten des Datentransfers (FTP, Heretrix) wurden ausgiebig getestet und verworfen.

- Das FTP-Verfahren ist besonders im Fall der Archivierung von Internetpräsenzen Dritter zu umständlich, erlaubt keine genaue Abgrenzung der zu archivierenden Daten und erforderte in den meisten Fällen eine aufwendige Rekonstruktionsarbeit.
- Spiegelungen mit dem Crawler Heretrix lieferten mangelhafte Ergebnisse besonders im Bereich von Bildern und Grafiken

Verwendete Software: Offline Explorer Pro, Teleport Pro, HTTrack, RM Recorder, Handy File

Bewertung

Bei der Archivierung von Websites muss die Bewertung als integraler Bestandteil der Erfassung organisiert werden. Eine nachträgliche Bewertung der gespiegelten Seiten ist wegen des hohen Arbeitsaufwandes nur in Ausnahmefällen möglich.

Anders als im konventionellen Bereich gehört die Festlegung von Intervallen, von Zeiträumen für den technischen Prozess der Spiegelung, die Bestimmung von Zeitfolgen und die Berücksichtigung bestimmter Anlässe mit zur Bewertung. Dies gilt auch für die Festlegung exemplarischer Spiegelungen ausgewählter Organisationsbereiche.

Zusammenfassung

Die Methode der Spiegelung als erster Schritt einer Archivierung von Webpräsenzen hat sich bewährt. Lediglich Datenbanken und einige passwortgeschützte Bereiche sind von der Spiegelung ausgeschlossen, können aber in einigen Fällen durch ergänzende Methoden erfasst werden.

Ein schwer zu handhabendes Problem stellen dynamische generierte Seiten dar. Sie sind zwar nach der Umwandlung in browsergestützte Formate prinzipiell zu spiegeln, führen aber unter Umständen zu einem gewaltigen Datenaufkommen. Um den Spiegelungsprozess in solchen Fällen nicht unterbrechen zu müssen, würde schon die Implementierung eines ‚Frühwarnsystems‘ in den Off Line Browser eine große Hilfe sein. Mit dem Hersteller wird über die Lösung dieser Frage weiterverhandelt. Die fehlerhaft Darstellung umgewandelter Dateien konnte mit Hilfe unterschiedlicher Verfahren behoben werden.

-Erschließung

Dokumentation der Metadaten

Generell ist bei der Diskussion zwischen der sichernden und der erschließenden Funktion der Metadaten zu unterscheiden. Die gängigen Standards sind in der Regel Mischformen mit unterschiedlichen Prioritäten. Der Anspruch Metadatensätze so anzulegen, dass sich mit ihrer Hilfe die Originale rekonstruieren ließen, ist in Hinblick auf die Datenmenge in den Spiegelungsprojekten illusorisch. Für die Arbeit am Internet-Archiv erwies sich die bisher vorherrschende Fokussierung auf die so genannten Metatags bei der Diskussion der Metadaten als eher hinderlich. Die Standardisierung der dokumentbezogenen Metatags war wesentlich im Hinblick auf das Einstellen wissenschaftlicher Publikationen ins Netz und deren bibliothekarische Erschließung entwickelt worden. Da wir aber nicht einzelne Dokumente archivieren, sondern ganze Internetpassagen, müssen innerhalb der Metadaten die projektbezogenen Erfassungsdaten von den dokumentbezogenen Erschließungsdaten, zu denen auch die Metatags gehören, unterschieden werden. Die Sicherung der Authentizität und Identität der archivierten Daten erfolgt wesentlich über eine Dokumentation der Erfassungsdaten.

Der beiliegende Metadatensatz (A1) befindet sich noch in der Diskussion. Die Semantik orientiert sich an METS, auch wenn der Standard (s.o.) nicht übernommen werden konnte. Als Syntax wurde ein Schema in XML gewählt. Auch wenn die große

Zahl dokumentbezogener Angaben (Fehler, Nachbearbeitungen, Sicherungsmaßnahmen) ein Problem darstellen, so erzielt man mit dieser Form eine

- Vereinheitlichung der unterschiedlichen Dokumentationen
- zusätzliche Formen der Präsentation (Internet)
- Möglichkeiten der gemeinschaftlichen Präsentation
- Entlastung der Verzeichnung von technischen Angaben
- Einbindungsmöglichkeiten in unterschiedliche Verfahren der Bereitstellung und Erschließung (Faust, Struktur) und einen
- leichteren Austausch der Informationen zwischen den Archiven

Außerhalb des im Abschnitt „Erfassung (erweiternde Methoden)“ genannten Datenbankverfahrens gibt es noch keine befriedigende Lösung für eine Automatisierung der Übernahme von Steuerungs- und Speicherdaten des Off Line Explorers. Es bleibt also zunächst noch bei dem von der Projektgruppe entwickelten Verfahren, das die entsprechenden Daten mit einem Javaskript ausliest und eine manuelle Nachbearbeitung erforderlich macht. Das Verfahren des Webarchivierungssystems muss, wie beschrieben, erst noch getestet werden.

Indexierung

Weiterhin bietet die automatische Volltextindexierung neben der Verzeichnung und den Metadaten vor allem mit den Möglichkeiten der diachronen und synchronen Suche eine wertvolle Ergänzung zu den Rechercheoptionen der übrigen Erschließungsformen. Versuche semantische Analyseverfahren in die Indexierung mit einzubeziehen konnten nicht fortgesetzt werden, weil die Handhabung der entsprechenden Software sich äußerst umständlich gestaltet und die Einbeziehung binärer Formate nicht gewährleistet werden konnte. Neben der Indexierungssoftware DtSearch kommt bei einigen Archiven copernic desktop zur Anwendung, vor allem um unterschiedliche Formate einer Spiegelung zusammenzufassen und anzeigen zu lassen. Die Recherchemöglichkeiten von copernic desktop werden besonders im ADL intensiv genutzt.

Eine Indexierungssoftware sollte folgenden Anforderungen entsprechen:

1. Freie Indexierung
2. Thesaurus (Optional)
3. Verarbeitung einer Datenmenge von mindestens 10 GB
4. Sprachmodul der deutschen Sprache/Stammformensuche
5. Boolesche Operatoren/Trunkierungen
6. Webform
7. gewichtete Anzeigen
8. Highlighting
9. browserfähige Ergebnisseiten
10. Ermöglichung von diachroner und synchroner Suche

Verzeichnung

Da in Deutschland generell verbindliche Standards für die archivarische Erschließung fehlen und die Praxis in den einzelnen Archiven von den jeweiligen Verzeichnungsrichtlinien geprägt ist, wurde innerhalb des Projekts auf die Festlegung verbindliche Standards verzichtet. Allerdings sind verbindliche Eingabefelder für die Erfassungsmasken festgelegt worden.

Eingabefelder in der Erfassungsmaske zur Verzeichnung von archivierten Websites:

- A. obligatorisch
 - allgemein
 - Bestand / Sammlung
 - Titel

- Signatur
- projektspezifisch
 - Link zur Seite (+ Adresse)
 - Metadaten (tech) (+Adresse)
 - Spiegelungsdatum / -zeitraum
- B. fakultativ
 - allgemein
 - Rechtliches
 - enthält
 - projektspezifisch
 - Anlass
 - (Sicherungskopie -falls vorhanden)

Generell herrschte Einvernehmen darüber, dass man eine Bevormundung der Benutzer durch ein übermäßiges Angebot an Verlinkungen und Verweisen vermeiden sollte. Was natürlich nicht ausschließt, dass es sich in einigen Fällen als durchaus sinnvoll erweisen kann, den Kontext durch Links zu anderen Spiegelungen oder zu in konventioneller Form vorliegendem Archivgut zu verdeutlichen. Die Datenbank Faust bietet jedenfalls die Möglichkeit entsprechende Felder einzurichten.

In einem halbautomatischen Verfahren können zudem auch gleichförmige Mengen digitalen Archivguts (Reden, Pressedienste) in einem Stapelverfahren von der Datenbank verarbeitet werden. Allerdings müssen die entsprechenden Stapel vorher angelegt werden. Besonders intensiv wird dieses Verfahren vom ACSP genutzt.

Zitierweisen:

Die Angaben zu Quellenbelegen aus dem Internet-Archiv müssen verständlich, eindeutig und dokumentspezifisch sein. Diese Anforderungen erfüllen sie nur in der Kombination aus Angaben zu Ort / Sammlung / Bestandsgruppe / Bestand / Titel (kurz) / Signatur / URL (entsprechend der Tektonik des jeweiligen Archivs). Im Fall des AdsD hätte die Belegstelle folgende Identifikation:

AdsD, Internet-Archiv, Internetseiten der SPD, SPD-Parteivorstand, Spiegelung vom 10.06.2002, Signatur 6/IntAr 00007,

URL:http://intar.fes.de/IntAr/SPD_B_P_2002_06_10/www.spd.de/servlet/PB/-s/In6rxt1evtbuxf8dkvkfosalzw5b86a/menu/1009319/index.html

Zusammenfassung

Die drei im Projekt entwickelten bzw. modifizierten Erschließungsformen können in Kombination miteinander oder einzeln eingesetzt werden. Jede Form bietet einen vollwertigen Zugang zum Internet-Archiv. In der Kombination ergibt sich die Möglichkeit, die einzelnen Erschließungsarten auf ihre jeweiligen Stärken zu beschränken. Im Idealfall könnte sich der Benutzer anhand der Metadaten einen Überblick über das Internetarchiv verschaffen, mit Hilfe der Verzeichnung einzelne Spiegelungen auswählen, die dann über die Indices durchsucht werden können.

Ein Desiderat bildet sicher noch die angemessene Einbeziehung nicht-diskursiver Bestandteile in die Erschließung. Hier müssen Ansätze, die etwa in der Verzeichnung auch ikonographische Beschreibungen mit einbeziehen, weiterverfolgt und ergänzt werden. So wird etwa vom ACSP ein Screenshot der Startseite bzw. der Sitemap in Faust aufgenommen.

Verwendete Software: Excel, dtSearch Web, Faust, XML-Spy

-Langzeitsicherung

Bei der Archivierung von Websites steht neben dem Erhalt des Content die Bewahrung der Funktionalität im Vordergrund der Langzeitsicherung. Die Konversion in ein

Archivformat kann also nur ergänzenden Charakter haben, da sie in der Regel mit dem Verlust der Funktionalität erkaufte würde. Ins Zentrum der Bemühungen muss also der Erhalt der Präsentationsformate gestellt werden. Zu den Besonderheiten der Archivierung von Webseiten gehört, dass nicht nur die große Zahl der unterschiedlichen Datei-Formate, die im Web Verwendung finden, von den Archiven nicht zu beeinflussen ist, sondern auch deren unterschiedliche, manchmal eigenwillige Programmierung. Dabei hat sich gezeigt, dass die Vielzahl der Anbieter mit den unterschiedlichsten Formen konventionswidriger Programmierungen der einzelnen Seiten eine automatische Konvertierung der gespiegelten Seiten in XHTML nicht zulässt.

Bis die verschiedenen Initiativen (Kopal, Nestor) zur Entwicklung konsistenter Langzeitarchivierungsstrategien geführt haben werden, die sich auch in unsere Projekte integrieren lassen, wird man sich mit einer Kombination einzelner Maßnahmen begnügen müssen, deren vorläufiger Charakter nicht geleugnet werden soll, die sich aber bewährt haben.

Speichermedien

- Sicherung auf einer Festplatte mit einem Raid-System
- Sicherung auf Bändern / Lagerung eines zweiten Bandes an einem anderen Ort
- Externe Sicherung auf CD oder DVD
- Sicherung über ein parallel geschaltetes Laufwerk

Anzustreben ist eine möglichst große Diversifikation der eingesetzten Medien zur Langzeitsicherung

Methoden

- Pflege eines Browserarchivs
- Dokumentation der Maßnahmen zur Langzeitsicherung in den entsprechenden Metadaten
- Protokollierung der Umgebungsdaten
- Statistik der zunehmenden bzw. abnehmenden Häufigkeit, mit der bestimmte Formate Verwendung finden
- Registrierung der Veränderung von Browser- bzw. Softwareunterstützungen
- Zusätzliche Sicherung in komprimierter Form. (Bei der Komprimierung mit WinZip entsteht, jedenfalls in einem für uns meßbaren Bereich, kein Datenverlust. Allerdings müssen bestimmte Parameter beachtet werden, um die Struktur der Dateien zu erhalten.)

Formate

- Da es keine verbindlichen Standards gibt, wird man sich im Fall der Ersetzung eines Formats an der Häufigkeit, mit der bestimmte Formate verwendet werden, orientieren müssen. Grundsätzlich ist darauf zu achten, dass bei Konversionen auch die entsprechenden Links umgeschrieben werden müssen.

Verwendete Software: WinZip, Nero, TapeWare

- Bereitstellung / Präsentationsformen

Die Bereitstellung der archivierten Websites erfolgt in allen Archiven im jeweiligen Intranet. Standard bei allen Archiven ist die Anbindung des Internet-Archivs an die Datenbank Faust. Schon jetzt wird dieser Zugang in einigen Archiven ergänzt durch die Bereitstellung von Suchmöglichkeiten über einen Index. Die Teststellung des ‚Off Line Web Archivs‘ erfolgt auch unter dem Aspekt der Schaffung eines weiteren benutzerfreundlichen Zugangs zum Internet - Archiv. Um die Präsenz des Projekts in der Öffentlichkeit zu optimieren, wurde damit begonnen, ausgewählte Metadaten zu den einzelnen Spiegelungen ins Internet zu stellen, so dass potentielle Interessenten

sich auch außerhalb des Intranets einen Überblick über die Bestände des Internet - Archivs verschaffen können.

- **Rechtliche Fragen**

Der Vorsitzenden des Arbeitskreises, D. Haak (ACDP) hat anlässlich der Projektvorstellung auf der Frühjahrstagung der Medienarchive erneut einen Überblick über die mit der Webarchivierung verbundenen rechtlichen Fragen gegeben. Die Unübersichtlichkeit der rechtlichen Situation rührt im wesentlichen daher, dass einige gesetzliche Regelungen zwar angekündigt, aber noch nicht verabschiedet wurden. Es lässt sich aber bereits jetzt die Tendenz erkennen, dass einige Bestimmungen des Copyrights für den digitalen Bereich sehr viel schärfer gefasst werden als für den konventionellen. Ein besonderes Problem stellen die Verwertungsrechte dar, die von der Betreibern der zu archivierenden Seiten mit einer zeitlichen Begrenzung erworbenen wurden.

2.4 Eigenleistung

Anschaffung der erforderlichen Hard- und Software durch die beteiligten Archive / Personalkosten für den Projektkoordinator.

2.5 Besonderheiten des Projektes:

2.6 Ergebnisse:

Die Projekte, die bereits über die jeweiligen Datenbanken erschlossen wurden, stehen auch schon für die Archivbenutzer bereit. Aber es ist immer noch zu früh, um Benutzerzahlen vorlegen zu können, da wir bisher nur einen relativ kurzen Zeitraum quellenmäßig abdecken können.

Die Webseiten des Projekts mit detaillierten Informationen zum Projekt, zu seiner Konzeption, den einzelnen Arbeitsschritten und der verwendeten Software sowie Hilfestellungen zu bestimmten Problemen zählte im September 2006 215 Benutzer in 199 Sitzungen.

Die Auswertung der archivierten Websites zum Nachweis elektronischer Zeitschriften, besonders von SPD-Ortsvereinszeitschriften, wird vom AdsD in Kooperation mit der Bibliothek der Friedrich-Ebert-Stiftung weitergeführt. Das Internet-Archiv hat sich auch bei der Erschließung von Bildbeständen als wertvolle Hilfe erwiesen, da mit diesen Informationen Bilder etwa von Parteitagern sehr viel schneller erschlossen und der Benutzung zur Verfügung gestellt werden können.

3. Öffentlichkeitsarbeit

Das Projekt wurde bisher auf insgesamt sieben nationalen und internationalen Veranstaltungen vorgestellt.

Zusätzlich zu den im letzten Zwischenbericht erwähnten Veranstaltungen wurde die Ergebnisse des Projekts auch auf dem

- Treffen des Archivnetzwerkes Bonn am 24. August 2006 in der Friedrich-Ebert-Stiftung und der
- XXXVIIth Annual Conference of the IALHI, Zürich, September 6-9, 2006 präsentiert.

Im Projektzeitraum wurden fünf interne und ein öffentlicher Workshop durchgeführt.

Das Projekt ist im Internet in insgesamt vier Themenportalen zur Webarchivierung und zwei weiteren Linksammlungen eingebunden.
(http://www.fes.de/archiv/spiegelung/projekt/links_projekte.htm)

3.1 Veröffentlichungen

Neben zahlreichen Pressemitteilungen wurden über die Arbeitsergebnisse des Projekts in insgesamt 24 Publikationen berichtet, darunter 18 Netzpublikationen. Fünf weitere Veröffentlichungen in konventioneller Form sind in Vorbereitung.

3.2 Ergebnisse der Öffentlichkeitsarbeit

In diesem Jahr liegen zur Vorstellung des Projekts noch Einladungen zum

- SPP/ICA: Annual Meeting 2.-4.11.2006 in Bonn und
- Deutsch-Niederländischen Archivsymposion 16. und 17. November 2006 vor

Zu den erfreulichen Ergebnissen des Projekts gehört auch, dass neben einzelnen besonders interessierten Archiven und neben den entsprechenden Ausschüsse der Bundeskonferenz der Kommunalarchive, die sich vor Ort über die Arbeit des Projekts informierten, auch die Archive des Siemens Konzerns und der HypoVereinsbank/ UniCredit Group auf Basis unserer Arbeitsergebnisse seit neuestem die Seiten ihrer Konzerne und Tochterunternehmen sowie einzelne Länderseiten sichern.

4. **Weitere Arbeiten und Planungen**

Projektplanung:

Der Ausbau der jeweiligen Internet-Archive wird vereinbarungsgemäß fortgesetzt. Die entsprechenden Kooperationsformen sind auf dem letzten Workshop vereinbart worden. (A2)

Rudolf Schmitz

Bonn, den 29.9.2006