

## AUS DEM ACDP

### **Fragen der Sicherung und Erschließung der Internetpräsenz der Parteien aus der Sicht des Archivs für Christlich-Demokratische Politik. Ein Projektbericht.**

Von Michael Hansmann

Das Internet, oder genauer gesagt ein Teil davon, das World Wide Web (WWW), ist ein junges, allerdings sehr flüchtiges Medium, das nicht mehr aus dem Computeralltag wegzudenken ist. Es hat sich zu einem universellen Kommunikations- und Informationsträger entwickelt und wird auch von Regierungen und Behörden aller Ebenen immer intensiver genutzt. Ebenso haben sich die politischen Parteien in Deutschland das neue Medium für die politische Arbeit und zur Verbreitung von Informationen zu Nutze gemacht.

Das Internet befindet sich im ständigen Fluss. So erscheinen monatlich Millionen neuer Internetseiten, während gleichzeitig andere unwiderruflich verloren gehen. Die durchschnittliche Lebensdauer einer Internetseite wird heute auf etwa 75 Tagen geschätzt und lässt Historiker zu dem Schluss kommen, dass große Lücken in der Überlieferung und damit der Geschichte zu befürchten sind, da unser digitales Erbe verschwindet.<sup>1</sup>

In diesem Artikel sollen die Hintergründe erläutert werden, warum eine Beschäftigung mit Internetauftritten nötig ist und wie das Internet den Gestaltungsraum der politischen Parteien erweitert hat. Dies geschieht vor dem Hintergrund der Bundestagswahl 2005 mit Schwerpunkt auf den Seiten der CDU. Es soll gezeigt werden, wie sich gerade das Archiv für Christlich-Demokratische Politik (ACDP) der Konrad-Adenauer-Stiftung mit der Archivierung digitaler Daten, in diesem Falle mit der Archivierung von Internetseiten, befasst und dabei neue Wege beschreitet. Dazu werden einige technische Erläuterungen gegeben, die für Verständnis im Umgang mit einer neuen Quellengattung werben.

Ein Teil des Öffentlichkeits- und Politikauftritts der Parteien wird mittlerweile nur noch im Internet veröffentlicht. Eine Sicherung des Internetauftritts der Parteien und damit auch eines Teils der digitalen Überlieferung ist dringend geboten, da ansonsten eine kaum behebbare Überlieferungslücke entstehen würde. Zur Relevanz der Sicherung der digitalen Überlieferung und der Furcht der Wissenschaft vor einem erneuten „Dunklen Zeitalter“ ohne oder mit wenigen Quellen sei auf das Werk von Bryan P. Bergeron verwiesen.<sup>2</sup>

---

1 Mikel PLETT, *Bewahrung digitaler Kultur. Vorschläge und Strategien zur Webarchivierung, eine neue Herausforderung nicht nur für Nationalbibliotheken*, Hamburg 2008.

2 Bryan P. BERGERON, *Dark Ages II. When the Digital Data Die*, Upper Saddle River, NJ 2002.

Für die Frühzeit des Internet und der digitalen Überlieferung hat sich das „Dunkle Zeitalter“ leider schon bewahrheitet. So sind die ersten Internetauftritte nicht mehr erhalten. Eine teilweise Ersatzüberlieferung existiert durch das sogenannte Internet-Archiv mit der „Wayback-Maschine“.<sup>3</sup> 1996 gründete Brewster Kahle das „Internet Archive“ und machte sich die Langzeitarchivierung digitaler Daten in frei zugänglicher Form zur Aufgabe.<sup>4</sup> Leider handelt es sich dabei um ein Projekt, auf das der Archivar aus Deutschland keinen Einfluss nehmen und die Sicherung und Aufbewahrung nicht aktiv mitgestalten kann. Auch weitere Nationalarchive und Nationalbibliotheken haben sich mit dem Thema der Archivierung von Netzpublikationen und Internetseiten auseinandergesetzt; hier sind die skandinavischen Staaten, Australien, Großbritannien und die USA zu nennen.<sup>5</sup>

Zum Thema Langzeitarchivierung digitaler Daten sei auch auf das Nestor-Projekt<sup>6</sup> verwiesen. Umfangreiche Informationen darüber können im Internet abgerufen werden. Auf der Web-Präsenz von Nestor sind auch viele Publikationen als Download erhältlich. Ein Gespräch am Stand des Nestor-Projekts anlässlich des Deutschen Archivtages 2006 zeigte leider deutlich, dass an eine Archivierung von Internetpräsenzen noch nicht gedacht wurde. Ein weiteres Projekt im Rahmen der digitalen Langzeitarchivierung heißt „Kopal“, dessen Ziel laut eigener Aussage ist, „den Aufbau einer technischen und organisatorischen Lösung um die Langzeitverfügbarkeit elektronischer Publikationen zu sichern“.<sup>7</sup> Keines der genannten Projekte konnte bis 2006 praktische Lösungen zur Archivierung von Internetseiten vorweisen.

---

3 <<http://www.archive.org/>> (Stand: 1.7.2008).

4 Aus: Wikipedia: <[http://de.wikipedia.org/w/index.php?title=Internet\\_Archive&stableid=46391293](http://de.wikipedia.org/w/index.php?title=Internet_Archive&stableid=46391293)> (Stand: 1.7.2008).

5 Als Literaturhinweise zum Thema Webarchivierung/Internetarchivierung seien als Beispiele genannt: Julien MASSANÈS, *Web Archiving*, Berlin 2006; Adrian BROWN, *Archiving Websites. A practical guide for Information Management Professionals*, London 2006; Antje SCHEIDING (SCHLIETER), *Archiving Websites. Archivierungskonzept für das Intranet der Dresdner Bank*, in: *VdA – Mitteilungen der Fachgruppe 6*, Nr. 29, S. 21–41, Sankt Augustin 2004; Frank TESKE, *Archivierung des Internets. Eine Aufgabe für Archive*, Marburg 2003. Eine umfassende Bibliographie zum Thema Webarchivierung mit dem Stand von 2006 befindet sich auf der Seite der niederländischen Archivschule als PDF-Datei <<http://www.archiefschool.nl/docs/ehbw.pdf>> (Stand: 1.7.2008).

6 „In Nestor arbeiten Bibliotheken, Archive, Museen sowie führende Experten gemeinsam zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Quellen. Ziel dieses vom Bundesministerium für Bildung und Forschung geförderten Projektes ist der Aufbau einer dauerhaften Organisationsform für alle Belange der Langzeitarchivierung sowie nationale und internationale Abstimmungen und Aufgabenteilungen“, aus: Nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland <<http://www.langzeitarchivierung.de/>> (Stand: 1.7. 2008).

7 <[http://kopal.langzeitarchivierung.de/index\\_ziel.php.de](http://kopal.langzeitarchivierung.de/index_ziel.php.de)> (Stand: 1.7.2008).

Die Internetauftritte der Parteien, der Fraktionen in den Landtagen und Abgeordnetenhäusern, der Vereinigungen und Sonderorganisationen und nicht zuletzt die persönlichen Auftritte der Politikerinnen und Politikern im Internet sind zu einem wichtigen und unverzichtbaren Teil der Politikwirkung und Selbstdarstellung geworden. Dies gilt gerade für Wahlkampfzeiten. So äußerte sich der damalige Generalsekretär der CDU, Volker Kauder, am 3. August 2005 vor der Bundestagswahl 2005 zur Bedeutung des Online-Wahlkampfes. Mit Blick auf die Zugriffszahlen, versicherte Kauder, seien die elektronischen Medien aus einem modernen Wahlkampf nicht mehr wegzudenken. Allein im Juli 2005 habe die Homepage <[www.cdu.de](http://www.cdu.de)> 4,2 Mio. Pageviews registriert. Im umgekehrten Verhältnis zur Reichweite stehen dabei die Kosten: So macht der Online-Wahlkampf nur ein Prozent des CDU-Wahlkampfetats in Höhe von 18 Mio. Euro aus. Auch hier kann „weniger mehr sein“, betonte Kauder.<sup>8</sup>

Der Internetwahlkampf der CDU 2005 zeigt die Notwendigkeit einer kontinuierlichen Spiegelung in Wahlkampfzeiten. Mit seinem Beginn veränderte sich der Internetauftritt der CDU auf Bundesebene ([www.cdu.de](http://www.cdu.de)) schlagartig. Der gewohnten Eingangsseite wurde ein neues, auf den Wahlkampf zugeschnittenes Wahlkampfportal vorgeschaltet. Ebenfalls auf der Wahlkampfseite wurde ein Link zu einer weiteren Sonderseite angelegt, in der Aussagen der Regierung Schröder plakativ dargestellt wurden. Schon die Auswahl der Internetadresse zeigt die Möglichkeiten des Internetwahlkampfes bereits in kleinen Details auf: <[www.leere-versprechen.de](http://www.leere-versprechen.de)>. Ein zusätzliches Instrument, um direkt auf Aussagen des politischen Gegners reagieren zu können, wurde mit der Sonderseite <[www.wahlfakten.de](http://www.wahlfakten.de)> geschaffen, die ebenfalls von der Eingangsseite der Wahlplattform zu erreichen war. Als Anwendungsbeispiel nannte Volker Kauder den bevorstehenden Auftritt von Gerhard Schröder im ARD-Sommerinterview. Hier konnte die CDU noch während des laufenden Interviews auf Aussagen von Schröder mit Gegendarstellungen im Internet reagieren.

Eine weitere Besonderheit der CDU-Internetwahlkampfplattform waren die extra zur Bundestagswahl angefertigten Internetseiten für bestimmte Zielgruppen. Es wurden Internetseiten für Auslandsdeutsche, Aussiedler, Jugendliche, Frauen und Senioren angelegt. Die Seiten standen zumeist unter einem Motto wie: „Aussiedler für den Wechsel“<sup>9</sup>, besaßen eine eigenständige Internetadresse, waren aber fest auf den Seiten der Wahlkampfplattform der CDU verankert. Hier wurde um Aussiedler geworben, auf die stets aussiedlerfreundliche Politik der CDU verwiesen, und die Positionen der rot-grünen Regierung wurden kritisch beleuchtet. Begleitet wurden die Seiten mit Aussagen von

---

8 ACDP, 30-001-008: <G:\Internetspiegelung\Bundesebene\cdu\CDU\_2005\_August-03\_kauder-vorstellung-internetwahlkampf\www.cdu.de\portal\9078\_9142.htm>.

9 <<http://www.aussiedler.cdu.de/default.htm>> (Stand: 3.8.2005).

Aussiedlern zu ihren Beweggründen, die CDU zu wählen sowie Links auf weitere Internetangebote der CDU oder ihrer Vereinigungen. Die anderen Internetseiten mit Werbung um spezielle Zielgruppen besaßen einen ähnlichen Aufbau.

*Warum und wie können flüchtige Internetangebote für die Zukunft gesichert und archiviert werden?*

Als Zentralarchiv der christlichen Demokratie gehört es zu den Kernaufgaben des ACDP, die geschichtliche Entwicklung der Christlichen Demokratie zu dokumentieren und zu erforschen.<sup>10</sup> Dazu gehört auch die Sicherung und Erschließung der Internetauftritte der Partei, ihrer Gliederungen und Vereinigungen, um diese der Forschung zur Verfügung zu stellen.

Ausgehend von den Pionierarbeiten bei der Archivierung von Internetseiten durch Rudolf Schmitz aus dem Archiv der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung hat sich das ACDP mit den Archiven der anderen politischen Stiftungen<sup>11</sup> zu einem von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekt zur Erfassung, Erschließung und Sicherung von Websites politischer Parteien der Bundesrepublik Deutschland sowie ihrer Fraktionen in den Parlamenten zusammengefunden.

Ziel des Projektes war es, Möglichkeiten zu erkunden, die Webauftritte der Parteien, Fraktionen, Vereinigungen und der Abgeordneten dauerhaft für die Forschung zu sichern. Dies gilt für innerparteiliche Diskussionen, programmatische Debatten, die Auseinandersetzung mit dem politischen Gegner, Wahlkämpfe sowie für die Selbstpräsentation von Abgeordneten, die verstärkt nur noch im Internet stattfindet. Wesentliche Anstöße kamen dabei von der Fachgruppe 6 (Parlaments- und Parteistiftungen) des VdA.<sup>12</sup>

Nach Sichtung der spärlich vorhandenen Literatur zum Thema Webarchivierung sowie Information über andere laufende Projekte zur Thematik erarbeitete die Projektgruppe ad-hoc-Lösungen, um möglichst ohne größeren Zeitverlust die für die beteiligten Archive relevanten Internetauftritte zu sichern. Dies wurde beschleunigt durch die Ankündigung der Bundestagswahlen 2005.

<sup>10</sup> Zu den Aufgaben des ACDP siehe: <<http://www.kas.de/wf/de/42.7/>>.

<sup>11</sup> Archiv der sozialen Demokratie (AdsD), Archiv für Christlich-Soziale Politik (ACSP), Archiv des Liberalismus (AdL) und Archiv Grünes Gedächtnis.

<sup>12</sup> Zur Vorgeschichte des Projektes und den Beweggründen vgl. Rudolf SCHMITZ, *Archivierung von Internetseiten. Spiegelungsprojekt im Archiv der sozialen Demokratie (AdsD)*, in: „Der Archivar“, 55 (2002), S. 135f.; DERS., *Aufbau und Struktur eines Internet-Archivs. Die Archivierung von Internetauftritten der SPD und ihrer Fraktionen in den Parlamenten*, in: „Der Archivar“, 57 (2004), S. 318–320; DERS., *Die Archivierung von Internetauftritten der politischen Parteien und ihrer Fraktionen in den Parlamenten*, in: *VdA – Mitteilungen der Fachgruppe 6*, Nr. 29, Sankt Augustin 2004, S. 11–20; DERS., *Politisches Internet-Archiv*, EBD. Nr. 30, Sankt Augustin 2005.

*Womit hat es der Archivar nun zu tun, wenn wir über das Internet und die Sicherung von Internetauftritten sprechen?*

Laut Definition im Online-Lexikon Wikipedia handelt es sich dabei um ein „Dokument im World Wide Web, das mit einem Webbrowser von einem Webserver abgerufen werden kann“<sup>13</sup>. Das Web-Dokument kann sich aus Text, Film und Bildern sowie Audio- und Animationselementen zusammensetzen und enthält in der Regel Verweise auf andere Internetseiten, sogenannte Hyperlinks. Internetseiten können statisch vorliegen, das heißt aus bereits programmierten Seiten bestehen oder dynamisch erstellt sein. Dynamisch generierte Internetseiten werden erst beim Aufruf des Nutzers durch Programme in Skriptsprachen wie PHP und mit Hilfe einer Datenbank bzw. eines Webserver erzeugt und dann dem Nutzer im Internet zur Verfügung gestellt. Während in den Anfangsjahren des World Wide Web die meisten Seiten manuell mit Hilfe eines Texteditors in der Seitenbeschreibungssprache HTML erstellt wurden, geschieht dies aktuell bei vielen – gerade dynamisch generierten – Seiten mit einem sogenannten Content Management System (CMS), wobei der HTML-Code und weitere Elemente automatisch erzeugt werden. In diesen Programmen können ganze Internetauftritte mit Hilfe bereitgestellter Vorlagen (Templates) fast ohne Programmierkenntnisse erstellt werden. Allerdings setzen auch diese Programme teilweise hohe Fachkenntnisse in Grafik- und Design voraus. Auf dem Markt befinden sich inzwischen zahlreiche Programme wie zum Beispiel Joomla! und TYPO3, die sich quasi zu Standard-open-service-CMS-Systemen entwickelt haben. Zu ihnen sind umfangreiche Handbücher teilweise auch online erhältlich.

Technisch bedingt gibt es bislang Grenzen bei der Erfassung und Sicherung von Internetseiten. Datenbanken lassen sich nicht spiegeln bzw. mit ihren Abfragemechanismen wieder zur Verfügung stellen. Im Bereich der Video- und Audiowiedergabe gibt es technische und rechtliche Grenzen bei den sogenannten Streaming-files, hier insbesondere beim Live-Streaming. So werden diese Dateien nicht mehr auf die Festplatte des Nutzers übertragen und von dort zur Ansicht abgespielt, sondern beim Aufruf eines Links nur noch online zur Verfügung gestellt, das heißt live beim Abruf übertragen.<sup>14</sup> Dies betrifft vor allem den Videobereich, bei dem es zu Problemen mit der rechtlichen Verwertung der Videodateien kommt. Um unbefugte Nutzung, wie das Kopieren von urheberrechtlich geschützten Materialien, zu unterbinden, versuchen die Anbieter der Video- bzw. Audiomaterialien, das Speichern auf lokalen Festplatten beim Nutzer zu verhindern. Dies geschieht durch das Live-Streaming bzw. durch

---

<sup>13</sup> Vgl. Wikipedia: <<http://de.wikipedia.org/w/index.php?title=Webseite&stableid=47876975>>.

<sup>14</sup> Aus: Wikipedia: <[http://de.wikipedia.org/w/index.php?title=Streaming\\_Media&stableid=47794431](http://de.wikipedia.org/w/index.php?title=Streaming_Media&stableid=47794431)>.

das zusätzliche Einpacken in sogenannte Containerformate, wie das Flash-Format<sup>15</sup> der Firma Adobe. Hier bekommt auch der Archivar Probleme, denn was nicht auf einen festen Speicher zu übertragen ist, kann auch nicht gesichert werden.

*Wie lassen sich Daten so speichern, dass sie auch wieder angezeigt werden können?*

Es gibt mehrere Verfahren,<sup>16</sup> von denen sich für die Zwecke des ACDP und auch der anderen Archive der politischen Parteien das sogenannte Spiegeln bzw. Webharvesting als die praktikabelste Lösung durchgesetzt hat. Bei dieser Methode wird eine Internetpräsenz aktiv erfasst, so wie sie ein normaler Nutzer zu sehen bekommt. Mit Hilfe eines Web-Crawlers, eines Programms zur Analyse und Sammlung von Informationen, wird der Inhalt einer zuvor im Programm eingestellten Internetadresse durchsucht und auf einem lokalen Datenträger abgelegt. Vor dem eigentlich technischen Vorgang des Spiegeln einer Internetpräsenz hat der Archivar allerdings noch einige Vorarbeiten zu leisten, u. a. die Bewertung der zu archivierenden Seite, denn selten genug besteht ein Internetauftritt aus nur einer Internetadresse. Oft sind auf den Seiten weitere Links zu Internetseiten integriert, die zumeist auf andere Internetadressen verweisen oder interne Unteradressen enthalten. Werden diese bei der Erfassung der Hauptadresse nicht berücksichtigt, bekommt man eine unvollständige Internetseite als Ergebnis. Gut kann man dies an der Internetpräsenz der CDU-Bundespartei, die unter der Adresse <www.cdu.de.> im Internet aufrufbar ist, erläutern. Um die Seiten sinnvoll zu archivieren und später wieder im Zusammenhang anzeigen zu können, müssen weitere Internetadressen, URL genannt, berücksichtigt werden. So befinden sich aktuell (Stand: 1. Juli 2008) mehrere URL auf der Homepage der CDU. Zu nennen wäre hier das Banner „Jetzt die Mitte stärken – Mitglied werden“, das den oberen Rand der CDU-Seite zentral beherrscht. Durch einen Klick auf dieses Banner wird der Nutzer auf die Mitgliederwerbeseite der CDU <www.mitglied-werden.cdu.de/> geführt. Ein weiterer Link, der scheinbar auf eine externe Seite führt, befindet sich im rechten Bereich der CDU-Seite. Es handelt sich hierbei um eine Portalseite zum Grundsatzprogramm der CDU mit der Adresse <http://www.grundsatzprogramm.cdu.de/>. Dieser Bereich verfügt ebenfalls über eine eigene URL, die den Benutzer vermeintlich von der CDU-Seite wegführt. Auf der Internetpräsenz der CDU gibt es weitere Beispiele für solche Portale bzw. Bereiche, die über andere Adressen verfügen, ohne die aber die CDU-Seite nicht vollständig

---

15 In einer Flash-Datei (Dateiendungen: flv oder swf) können Text-, Bild- und Tondokumente zusammengefasst werden und über eine einzige Datei ausgegeben werden.

16 Vgl. Anm. 5.

wäre. Diese Adressen müssen ebenfalls dem Erfassungsprogramm mitgeteilt werden. Andere Links wiederum werden bewusst im Rahmen der Bewertung der Internetseite nicht zur Spiegelung ausgewählt, da sie eigenständige Seiten sind bzw. nach dem Provenienzprinzip separat erfasst werden. Zu nennen sind hier die Seiten mit Links auf die Landesverbände der CDU, die Kreisverbände, die Vereinigungen und Sonderorganisationen der CDU und die Internetpräsenz der CDU/CSU-Fraktion im Deutschen Bundestag.

Während der DFG-Projektphase nutzten die beteiligten Archive den Web-Crawler „OfflineExplorer“<sup>17</sup>, ein kommerzielles Programm. Mit Hilfe dieses Spezialprogramms können Internetpräsenzen auf eine Festplatte oder einen Server im lokalen Netzwerk des ACDP gespeichert werden. Ein gravierender Nachteil dieses Programms ist, dass bei jedem Spiegeln einer Internetpräsenz jeweils die Internetseite komplett neu gespeichert und abgelegt wurde. Dies führte zu einem rasanten Anwachsen der Datenmengen und dem Bedarf nach immer größerem Speicherplatz auf den Servern. Im Durchschnitt wurden bei jeder Spiegelung des Internetauftritts der Bundes-CDU ca. 950 MB an Daten gespeichert. Die Datensicherung in der IT-Abteilung der Konrad-Adenauer-Stiftung wurde mit der schnell in den Gigabytebereich wachsenden Datenmenge vor das Problem gestellt, die tägliche Datensicherung auf Dauer zu gewährleisten. So drängte die IT-Abteilung auf eine integrierte Lösung, bei der die Daten der Internetspiegelung von den übrigen Daten des Hauses getrennt werden sollten.

Ein weiterer Kritikpunkt an dem funktionierenden, allerdings nicht komfortablen und effizienten Verfahren ergab sich aus dem Umstand, dass die Spiegelungen lokal von einem zweiten Arbeitsplatzrechner aus gesteuert wurden. Auf diesem Standard-PC befanden sich neben der hausüblichen Software noch die zur Archivierung benötigten Programme, u.a. auch der bereits erwähnte OfflineExplorer. Stürzte der PC im Betrieb ab, waren auch die gerade laufenden Projekte verloren. Größtes Manko war allerdings die Tatsache, dass aus Sicherheitsgründen der PC nach Dienstschluss ausgeschaltet werden musste. Damit konnten in den Nachtstunden, in denen eine Störung des Internetzugangs durch den Abruf großer Datenmengen niemanden in der KAS belastete, keine Spiegelungen durchgeführt werden.

Das DFG-Projekt zur Internetarchivierung war auf zwei Jahre angelegt und konnte erfolgreich Ende 2006 mit dem Nachweis, dass eine weitgehende Sicherung und Archivierung von Internetseiten mit der aktuellen Technik möglich ist, abgeschlossen werden.<sup>18</sup> Es wurden Lösungen gefunden, die allerdings

---

17 OfflineExplorer Pro der Firma Metaproducts in der Version 3, Informationen über die weiterentwickelte Version unter <<http://www.metaproducts.com/>> (Stand: 1.7.2008).

18 Projektseite unter <<http://www.fes.de/archiv/spiegelung/default.htm>> (Stand: 1.7.2008).

sehr zeitaufwendig und technisch schwierig und in konzeptioneller Sicht nicht befriedigend waren.

Nach Projektabschluss wurde weiter gespiegelt und im ACDP entschieden, das während der Projektphase erworbene Wissen zu vertiefen und die Technik zur Archivierung weiterzuentwickeln. Auf der Grundlage des bisher Erreichten und der Definition der grundsätzlichen Ansprüche an ein neues System zur Erfassung, Archivierung und Wiederanzeige von Internetseiten wurde in Kooperation mit einer privaten Softwarefirma eine datenbankgestützte Lösung zur redundanzfreien Archivierung von Internetauftritten realisiert.

Entscheidend für die Zukunft war es in diesem Zusammenhang, eine OAIS-konforme Lösung zu erreichen, um keine proprietäre Insellösung zu schaffen. Bei OAIS, der Abkürzung für Open Archival Information System<sup>19</sup>, handelt es sich um ein ISO-zertifiziertes Referenzmodell, das u. a. von der NASA entwickelt wurde. Dies geschah vor dem Hintergrund, dass elektronische Dokumente auf längere Sicht nicht mehr zur Verfügung stehen. Beim OAIS-Funktionsmodell werden u. a. sechs Aufgabenbereiche genannt: Die Übernahme der Daten (Ingest), die Aufbewahrung (Archival storage), das Daten Management, die Systemverwaltung, die vorausschauende digitale Bestandserhaltung (Preservation Planning) und der Zugang (Access) zu den Daten.<sup>20</sup>

#### *Das „Offline Web Archiv“*

In enger Kooperation mit der beauftragten Firma gelang es, mit dem „Offline Web Archiv“ (OWA)<sup>21</sup> eine marktreife Lösung zu entwickeln, die OAIS-konform ist.<sup>22</sup> Mit dieser Hardware/Software-Lösung wurde ein Instrument zur Langzeitsicherung<sup>23</sup> digitaler Daten aus dem Internet geschaffen. Die Entwicklung des Programms begann Ende 2006 und konnte 2007 den Status der Einsatzbereitschaft erreichen. Parallel zur Entwicklung des Programms begann

19 Das Papier als PDF-Datei unter der Adresse: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>> (Stand: 1.7.2008).

20 Eine sehr gute deutschsprachige Erläuterung bietet Nils BRÜBACH, *OAIS – Das „Open Archival Information System“*. Ein Referenzmodell zur Organisation und Abwicklung der Archivierung digitaler Unterlagen, Seite 6. Der Artikel steht online unter der Adresse: <[http://www.archiv.sachsen.de/download/pp\\_bruebach.pdf](http://www.archiv.sachsen.de/download/pp_bruebach.pdf)> (Stand: 1.7.2008).

21 <<http://www.oia-duesseldorf.de/>> (Stand: 1.7.2008).

22 „Mit dem Offline Web Archiv (OWA) hat die oia Dr. Salm ein OAIS-konformes System speziell für die digitale Archivierung von Webpräsenzen entwickelt“, aus der Produktbeschreibung der Firma oia: <<http://www.oia-duesseldorf.de/default-Dateien/Page472.htm>> (Stand 1.7.2008).

23 Der Begriff Langzeitarchivierung ist eine Tautologie – so wie ein „Weißer Schimmel“ –, hat sich aber leider in der Fachwelt eingebürgert. Eine Archivierung sollte eigentlich immer eine unbegrenzte Sicherung darstellen.



im ACDP die Begleitung der Entwicklung und die Testphase der Programmfunktionen bis zum Ende der ersten Entwicklungsphase.

Beim OWA handelt es sich um eine erweiterbare Client-Server-Lösung, die es ermöglicht, den Server über einen Client, in diesem Falle den normalen Büro-PC zu steuern. So ist es nun dank der implementierten Softwarearchitektur möglich, den Client nach der Einrichtung der geplanten Spiegelungen, die verstärkt in den Nachtstunden stattfinden, auszuschalten, während alle Arbeitsaufträge durch den im Rechenzentrum stehenden Server abgearbeitet werden. OWA ermöglicht ein redundanzfreies Spiegeln von Internetseiten und vermeidet durch den im Programm integrierten DataAnalyzer, dass bei jedem Spiegelungsvorgang erneut alle Daten der Internetseiten einer Internetpräsenz heruntergeladen, sondern nur neue oder geänderte Dateien eines Internetauftritts in die Datenbank überführt werden. Dadurch werden erheblich weniger Daten, die unnötig Speicherplatz verbrauchen würden und letztlich auch eine Kostenfrage sind, auf den Servern abgelegt. Durch OWA lassen sich die Internetseiten wieder so zu einer Internetpräsenz herstellen, wie sie am Tage der Archivierung ausgesehen haben.

Bevor eine endgültige Archivierung geschieht, erfolgt eine Bearbeitung der Spiegelungen durch den Archivar. Dies geschieht mit Hilfe von OWA auf dem Client-PC. Die Spiegelungen können auf Vollständigkeit und Funktionsfähigkeit überprüft und bei Bedarf fehlende Teile mittels eines Online-Abgleichs nacherfasst werden. Erst nach Abschluss dieser Bearbeitung erfolgt die endgültige Archivierung einer Internetpräsenz, und damit sind weitere Arbeiten an der Seite nicht mehr möglich.

OWA hält eine Vielzahl von Metadaten über die Spiegelungen bis hin zu Auswertungen bereit. Diese Metadaten sind für den Betrieb und die Nutzung des Offline Web Archivs erforderlich und werden im XML-Format zur Ausgabe bereitgestellt. So werden alle technischen Einstellungen, die für eine Spiegelung getroffen werden, dokumentiert. Es ist auch möglich, eine Spiegelungshistorie darzustellen, inklusive der Auflistung aller vorhanden Dateitypen sowie die Datengröße der einzelnen Spiegelungen.<sup>24</sup>

Ein Modul zum Export der archivierten Spiegelungen ist implementiert und damit die Betrachtung und mögliche Weiterverarbeitung der exportierten Internetpräsenz oder einzelner Daten ohne OWA möglich. Dies ist für die Zukunftsplanung ein entscheidender Faktor, denn sollte der Fall eines Systemwechsels eintreten, wären die Daten weiter zugänglich.

---

24 „Sämtliche Parameter, die zur Speicherung einer Webpräsenz geführt haben, die Aktionen der Benutzer und des Systems, werden protokolliert und im Bestand festgehalten, so dass sämtliche Metadaten einer archivierten Spiegelung zur Interpretation der Überlieferungen mit herangezogen werden können“, <<http://www.oia-duesseldorf.de/default-Dateien/Page472.htm>> (Stand: 1.7.2008).

Für die digitale Langzeitarchivierung ist ein Punkt besonders hervorzuheben. Wie im OWA-Modell gefordert, ist der Punkt der vorausschauenden digitalen Bestandserhaltung (Preservation Planning) bereits konzeptionell in OWA vorgesehen, hier Erhaltungsplanung genannt. Eines der Hauptprobleme der digitalen Archivierung ist das Veralten von Dateitypen oder Probleme mit der Versionierung, was letztlich zur Unlesbarkeit der Daten führen kann. Man denke nur an Word-Dokumente der ersten Versionen, nicht mehr aktuelle Textverarbeitungsprogramme oder Bilddateien, deren Format heute nur noch selten vorkommt.

Das OWA stellt eine Funktion zur Migration der Daten zur Verfügung. Damit ist es möglich, Dateien gezielt nach Dateityp und Version zu exportieren. Außerhalb von OWA kann eine Konvertierung in ein neues lesbares Format vorgenommen und anschließend die konvertierten Dateien wieder in OWA importiert werden. Die „Originaldateien“ bleiben zusätzlich erhalten. Auch diese Vorgänge werden vom System protokolliert und sind als Metadaten zur Bestandshistorie abrufbar.

Für die qualitative Nutzung der archivierten Internetpräsenzen ist eine sinnvolle Erschließung der neuen Quellengattung nötig. Die Erschließung und Verzeichnung erfolgt im Archiv für Christlich-Demokratische Politik mit Hilfe der Standardarchivdatenbank FAUST 6.<sup>25</sup> Durch die Nutzung der bewährten Standardsoftware FAUST stehen die archivierten Internetpräsenzen nach Abschluss der Erschließung den Kolleginnen und Kollegen und später dem Benutzer zur Verfügung.

Für die Einbindung in FAUST und die Bestandstektonik des ACDP wurde in der Beständesystematik<sup>26</sup> eine neue Bestandsgruppe innerhalb des Bereichs der audiovisuellen Medien geschaffen. So werden die archivierten Internetseiten der Bundes-CDU unter der Bestandssignatur 30-001 (Internetpräsenz der CDU Deutschlands) geführt, die Internetseiten der CDU/CSU-Fraktion besitzen die Signatur 30-002, während für die Internetauftritte der Politikerinnen und Politiker der CDU die Bestandsgruppe 30-007 reserviert wurde.<sup>27</sup>

OWA stellt eine Schnittstelle nach FAUST zur Verfügung, so dass einzelne archivierte Spiegelungen oder eine Gruppe von Spiegelungen, wie z.B alle

25 Informationen zu FAUST unter < <http://www.land-software.de/> > (Stand: 1.7.2008).

26 Günter BUCHSTAB (Hg.), *Die Bestände des Archivs für Christlich-Demokratische Politik der Konrad-Adenauer-Stiftung*, 5. Aufl., Sankt Augustin 2004, aktualisierte Fassung unter <<http://www.kas.de/wf/de/71.3772/>> (Stand: 1.7.2008).

27 Michael HANSMANN, *Erschließung archivierter Internetressourcen. Ergebnisse eines DFG-Projektes zur Sicherstellung der dauerhaften Zugänglichkeitmachung von Websites der Parteien aus der Sicht des Archivs für Christlich-Demokratische Politik*, in: Frank M. BISCHOFF (Hg.), *benutzerfreundlich – rationell – standardisiert. Aktuelle Anforderungen an archivische Erschließung und Findmittel* (Beiträge zum 11. Archivwissenschaftlichen Kolloquium der Archivschule Marburg), Marburg 2007, S. 203–218.

Spiegelungen der CDU-Bundespartei, durch OWA angezeigt werden. Dies geschieht in FAUST mittels eines Digitaldokumentenfeldes. Nach Auswahl des Feldes startet FAUST das Offline Web Archiv und stellt die gewünschte Internetpräsenz zur Verfügung.

*Sollen Internetspiegelungen verzeichnet werden?*

Natürlich stellt sich die Frage, ob es sinnvoll und möglich ist, eine archivierte Internetpräsenz zu erschließen und zu verzeichnen. Im ACDP wird dies bejaht.

Die wesentlichen Merkmale für eine Erschließung, auch nach der Norm ISAD (G)<sup>28</sup> sind gegeben. So kann ein Titel gebildet, ein „Enthält-Vermerk“ erstellt, und die Laufzeit angegeben werden. Auch eine Umfangsangabe ist möglich, in diesem Fall nicht durch eine Zentimeterangabe, sondern durch die Datengröße. Ebenso ist eine Verschlagwortung mittels Personenindex und Sachthesaurus möglich.

Bei der Erschließung und Verzeichnung muss der Archivar die Besonderheiten des Aufbaus und des Inhaltes der archivierten Internetseiten beachten, den Inhalt bewerten und schließlich eine Entscheidung zur Auswahl der zu beschreibenden Elemente treffen. Das ACDP hat sich entschieden, eine Verzeichnung sowohl nach inhaltlichen als auch ikonographischen Merkmalen vorzunehmen. Eine Internetseite lebt gerade von der grafischen Gestaltung, inklusive dem Einsatz von Bild, Ton und Filmelementen. Der Benutzer erhält als Ergebnis ein strukturiertes Findmittel, das sogar, wenn gewünscht, als konventionelles Findbuch entweder in Papierform oder als PDF-Datei zur Verfügung gestellt werden kann. Daneben stehen in OWA selber noch Volltextsuchfunktionen zur Verfügung, weitere Suchmöglichkeiten werden in der nächsten Version integriert. Den Erfolg der maßgeblich durch das ACDP beeinflussten Entwicklung kann man an der erfolgreichen Verbreitung dieser auf dem Markt noch einzigartigen Software belegen. So ist das Offline Web Archiv (OWA) mittlerweile bei der Deutschen Post AG, Siemens, dem Archiv der sozialen Demokratie der Friedrich-Ebert-Stiftung, dem Archiv für Christlich-Soziale Politik der Hanns-Seidel-Stiftung und anderen Firmen und Archiven im Einsatz.

Seit der Einführung des OWA Mitte Mai 2007 konnten 1.119 Spiegelungen durchgeführt und insgesamt 8.374.870 Dateien archiviert werden. Diese belegen 243 Gigabyte an Speicherplatz auf dem Archivierungsserver. Ohne die Möglichkeit zum redundanzfreien Speichern wären wir sicherlich schon lange

---

28 ISAD(G): *Internationale Grundsätze für die archivische Verzeichnung*, übersetzt und überarbeitet von Rainer Brüning, Werner Heegewaldt und Nils Brübach (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft; Nr. 23, Ed. 2), 2. Aufl., Marburg 2006.

im Terrabyte-Bereich angelangt, verbunden mit entsprechenden Hardwarekosten.

In erster Linie wurden und werden die Internetauftritte der Bundes-CDU sowie der Landesverbände der CDU im wöchentlichen Rhythmus gespiegelt. Dazu kommen monatlich die Seiten der Vereinigungen. Der nächste Schwerpunkt wird auf dem Bereich der Bezirks- und Kreisverbände der CDU und auf den persönlichen Internetauftritten führender CDU-Persönlichkeiten liegen.

Bedenken gibt es in rechtlicher Hinsicht.<sup>29</sup> Das Urheberrecht kann sich jetzt und auch zukünftig, da die Regelungen immer restriktiver werden, auf die Bereitstellung für Benutzer negativ auswirken. So ist es zurzeit nicht möglich, die archivierten Internetseiten einem breiten Publikum online zur Verfügung zu stellen. In erster Linie geht es hier um die Frage nach den Verwertungsrechten für Bilddateien, Audio- und Filmsequenzen. Rechtlich ist es möglich, Fotos nur für zwei Jahre zur Nutzung zu erwerben. Dies führt zwangsläufig bei archivierten Seiten, die später auch gezeigt werden sollen, zu rechtlichen Problemen. Hier müssen die betroffenen Archive mit den Auftraggebern der Internetseiten zu Lösungen kommen. Bis zu einer Regelung wird sich die Nutzung auf den Lesesaal eines Archivs beschränken müssen.

---

<sup>29</sup> Zur Rechtsproblematik: Dietmar HAAK, *Rechtliche Fragen zur Websitearchivierung*, in: *VdA – Mitteilungen der Fachgruppe* 6, Nr. 31, Sankt Augustin 2006, S. 27–41.