

Archivierung des Internets - Eine Aufgabe für Archive?

Transferarbeit von

Frank Teske

**Eingereicht am Hauptstaatsarchiv Stuttgart und an der
Archivschule Marburg am 1. April 2003**

Inhaltsverzeichnis

ZUSAMMENFASSUNG.....	2
1 EINLEITUNG.....	3
2 AUSWAHL DER WEBSITES.....	4
3 BESCHREIBUNG UND BEWERTUNG DER WEB-SITES	6
3.1 DIE INTERNETSEITEN DER LANDESMINISTERIEN	6
3.2 DIE INTERNETSEITEN DES LfD UND DER FÜHRUNGSAKADEMIE	11
4 STRATEGIEN ZUR ARCHIVIERUNG VON ONLINE-PUBLIKATIONEN...14	
4.1 PROJEKTE IN DEUTSCHLAND.....	14
4.2 PROJEKTE IM AUSLAND	18
5 KONZEPT FÜR DIE ARCHIVIERUNG VON WEB-SITES DURCH DAS HSTAS.....	22
6 LITERATUR.....	28
7 ANHANG: ABKÜRZUNGSVERZEICHNIS.....	30

Zusammenfassung

Online-Publikationen sind nicht nur allgemein als wichtige historische Quellen einzustufen, sondern können im Einzelfall auch eine sinnvolle Ergänzung von Archivbeständen darstellen. Für eine Übernahme durch das Hauptstaatsarchiv Stuttgart (HSTAS) kommen vor allem die Websites der Ministerien des Landes Baden-Württemberg in Betracht, die nach eingehender Analyse alle als archivwürdig einzustufen waren. Vor der Erstellung eines Konzepts zur Archivierung dieser Internetseiten durch das HSTAS wurden zunächst einzelne Projekte beleuchtet, die auf verschiedene Weise das Ziel verfolgen, Netzpublikationen dauerhaft zu sichern. Die bisherigen Erfahrungen der Deutschen Bibliothek, des Bundesarchivs und anderer Stellen im In- und Ausland konnten abschließend für die Entwicklung eines Workflows für die Übernahme von Websites durch das HSTAS nutzbringend eingebracht werden.

Für die Betreuung der vorliegenden Arbeit und die Übernahme der Gutachten habe ich Herrn Dr. Nils Brübach (Archivschule Marburg) und Herrn Dr. Robert Kretschmar (HSTAS) zu danken.

Herrn Dr. Albrecht Ernst (HSTAS), Herrn Hans Liegmann (DDB) sowie Frau Dr. Andrea Süchting-Hänger (BArch) danke ich für die bereitwillige Erteilung von Auskünften. Darüber hinaus bin ich den Mitarbeiterinnen und Mitarbeitern der in die Untersuchung einbezogenen Dienststellen zu Dank verpflichtet, deren Befragung eine Grundlage für die Bewertung der einzelnen Internetpräsentationen darstellte.

1 Einleitung

Obwohl erkannt wurde, dass Online-Publikationen als Bestandteil des kulturellen Erbes, als wichtige Quellen für künftige Generationen und somit als zu bewahrendes Kulturgut anzusehen sind, ist die Frage der Archivierung von Websites bisher noch immer ungelöst.

Archive nutzen in zunehmender Weise die vielfältigen Möglichkeiten, die ihnen das Internet als Plattform der Öffentlichkeitsarbeit und zur Ausweitung ihrer Produktpalette bietet¹; ob die Archive jedoch auch Verantwortung dafür tragen, dass die Inhalte dieses seit Mitte der 90er Jahre rasant wachsenden Mediums dauerhaft erhalten bleiben, wird vorläufig kaum diskutiert.

Mit vorliegender Studie soll daher am Beispiel des HSTAS untersucht werden, welchen Beitrag Archive zur dauerhaften Sicherung von Internetseiten leisten können. Dabei ist zunächst zu eruieren, welche Inhalte des Internets möglicherweise für eine Archivierung in Frage kommen. Danach sollen ausgewählte Websites auf deren Archivwürdigkeit hin analysiert werden. Nach der archivischen Bewertung der Internetseiten soll dann ein möglicher Lösungsansatz für das Problem der Archivierung dieser besonderen Form digitaler Unterlagen vorgestellt werden. Insbesondere gilt es, methodisch-technische Fragen zu thematisieren; hierbei können verschiedene Projekte zur Archivierung von Online-Publikationen im In- und Ausland als Grundlage für die Entwicklung eines Workflows zur Übernahme von Websites durch das HSTAS dienen.

¹ Vgl. das Online-Angebot der Landesarchivdirektion Baden-Württemberg und der Staatsarchive des Landes unter <http://www.lad-bw.de>.

2 Auswahl der Websites

Bevor über Strategien zur Archivierung von Internetseiten nachgedacht wird, ist zunächst zu prüfen, welche Inhalte des Internets überhaupt relevant für eine dauerhafte Aufbewahrung erscheinen. Aus archivischer Sicht kommen dabei an erster Stelle jene Online-Publikationen in Betracht, die eine sinnvolle Ergänzung der Bestände des einzelnen Archivs darstellen könnten.²

Hinsichtlich der Bestände des HSTAS wäre dabei zunächst an die Websites jener Dienststellen zu denken, für deren Schriftgut das Hauptstaatsarchiv zuständig ist. Ablieferungspflichtige Dienststellen sind derzeit die zehn Landesministerien, die Landesvertretung beim Bund, der Staatsgerichtshof, der Landesbeauftragte für den Datenschutz, die Führungsakademie, der Landespersonalausschuss, die Landesrektorenkonferenz und - nicht zu vergessen - das HSTAS selbst.³

Der Landespersonalausschuss und die Landesrektorenkonferenz sind nicht im Internet vertreten, die Landesvertretung beim Bund ist der Website des Staatsministeriums⁴ angegliedert und Informationen zum Staatsgerichtshof schließlich sind über die vom Staatsministerium gepflegte Homepage der Landesregierung⁵ abzurufen.

Eine Bewertung des Web-Auftritts des HSTAS⁶ und die daran gegebenenfalls anschließende Entwicklung eines Konzepts zur Archivierung der Website kann sinnvoll nur im Zusammenhang mit einer archivischen Bewertung des gesamten Internetangebots der Staatlichen Archivverwaltung Baden-Württemberg⁷ erfolgen. Da dies im Rahmen dieser Studie nicht möglich ist, muss für folgende Überlegungen die Internetseite des Hauptstaatsarchivs unberücksichtigt bleiben.

Einer genaueren Betrachtung zu unterziehen sind hingegen die Websites aller

² Gemäß den *Richtlinien für die Sammlungstätigkeit in den baden-württembergischen Staatsarchiven* können die Staatsarchive archivwürdige Druckschriften (als solche wären Websites zu interpretieren), die von Dienststellen des Landes in ihrem Zuständigkeitsbereich herausgegeben werden und die Einblicke in die Aufgaben, Organisation und Arbeit der Behörden geben, als Sammlungsgut übernehmen (vgl. <http://www.lad-bw.de/lad/risam.htm>).

³ Auskunft von Herrn Dr. Albrecht Ernst (HSTAS) vom 17. Februar 2003.

⁴ URL: <http://www.stm.baden-wuerttemberg.de>.

⁵ URL: <http://www.baden-wuerttemberg.de>.

⁶ URL: <http://www.lad-bw.de/hstas/>.

⁷ URL: <http://www.lad-bw.de>.

Landesministerien, des Landesbeauftragten für den Datenschutz sowie der Führungsakademie. Mit einzuschließen ist zudem die Seite der Landesregierung, da dieses sogenannte *Landesportal* in den Aufgabenbereich des Staatsministeriums fällt und über diese Internetseite der Zugang zur Online-Präsentation des Staatsgerichtshofs erfolgt.

Nicht in die Untersuchung einbezogen werden konnte das behördeninterne *Intranet*. Eine Bewertung der Intranetseiten, die von der Hälfte der genannten Dienststellen mit eigener Website zusätzlich betrieben werden⁸, war neben der zeitaufwendigen Analyse der einzelnen Internetauftritte nicht zu leisten. Sicherlich ist für die Zukunft im Bereich der Ergänzungsüberlieferung auch an eine Archivierung des Intranets zu denken. Allerdings ist zu vermuten, dass die archivwürdigen Inhalte des Intranets (wie Organisationshandbücher, Arbeitsrichtlinien etc.) ohnehin größtenteils auch als Schriftgutüberlieferung ins Archiv gelangen.

⁸ Diese Intranetdienste sind nicht zu verwechseln mit dem baden-württembergischen Landesverwaltungsnetz, an das alle Dienststellen angeschlossen sind. Dem Landesverwaltungsnetz hat sich Katharina Ernst im Rahmen ihrer im April 2002 am Hauptstaatsarchiv Stuttgart und an der Archivschule Marburg eingereichten Transferarbeit *Die Möglichkeiten des Internets/Intranets als Mittel der Kooperation zwischen Archiven und Behörden* gewidmet.

3 Beschreibung und Bewertung der Websites

3.1 Die Internetseiten der Landesministerien

Die Einrichtung von Internetseiten durch die Ministerien des Landes Baden-Württemberg erfolgte im wesentlichen in den Jahren 1998-2000. Bereits 1996 war zunächst die Seite der Landesregierung ins Netz gestellt worden, ein Jahr später hatte der Ministerrat dann beschlossen, den Ministerien die Möglichkeit zum Aufbau einer eigenen Website zu geben.⁹

Bis Ende 1998 waren zunächst das Wirtschaftsministerium und die Ministerien für Umwelt und Verkehr, für Kultus, Jugend und Sport sowie für Ernährung und Ländlichen Raum im Internet präsent. Im Zusammenhang mit dem Aufbau der Website des Staatsministeriums, die im Oktober 1999 freigeschaltet wurde, entschloss sich die Landesregierung dazu, Gestaltungsvorgaben für die Internetauftritte zu formulieren, um auf eine möglichst einheitliche Präsentation der Ministerien im Internet hinzuwirken. Zu Beginn des Jahres 2000 wurde dementsprechend das Landesportal dem Design der Seite des Staatsministeriums angepasst. Auch das Innen-, das Finanz-, und das Justizministerium, die im Jahr 2000 ihre Internetseiten starteten, übernahmen dieses Modell. Lediglich das Sozialministerium und das Ministerium für Wissenschaft, Forschung und Kunst entschlossen sich zu einer anderen Gestaltung im Rahmen der Landesvorgaben; das Sozialministerium passte jedoch bereits im November 2001 seinen Internetauftritt dem Landesportal an.

Derzeit (März 2003) präsentieren sich demzufolge fünf Ministerien und die Landesregierung mit einheitlichem Design im Internet; die Websites der anderen fünf Ministerien unterscheiden sich hingegen in der Gestaltung zum Teil erheblich, abgesehen von wenigen Merkmalen, wie etwa dem stets im oberen linken Eck platzierten Landeswappen.

Die Präsentationen der **Landesregierung (Landesportal)** sowie des **Finanz-, In-**

⁹ Folgende Ausführungen basieren auf Gesprächen mit den für die Internetauftritte zuständigen Mitarbeitern der Ministerien, die Ende Februar und Anfang März 2003 geführt wurden.

nen-, Justiz-, Sozial- und Staatsministeriums im Internet¹⁰ erfolgen nach einheitlichem Schema:

Unter einer Titelleiste¹¹, die beim Blättern durch die jeweiligen Seiten unverändert bleibt, sind drei Breiche nebeneinander angeordnet; links ist das Menü zur Navigation durchs das Internetangebot untergebracht, in der Mitte wird die jeweils ausgewählte Information angezeigt und rechts befinden sich teilweise animierte Schaltflächen, hinter denen sich Verknüpfungen zu internen Seiten (z.B. die Grußworte der Minister) aber auch Links zu externen Websites (etwa zur Werbe- und Sympathiekampagne des Landes „Wir können alles. Außer Hochdeutsch.“¹²) verbergen. Bei der Auswahl eines Menüpunktes im linken Bereich werden die Buttons auf der rechten Seite durch eine Menüleiste ersetzt, über welche Detailinformationen zum ausgewählten Thema abgerufen werden können.

Inhaltlich bieten die Internetseiten vor allem Informationen über Aufbau und Organisation der einzelnen Ministerien sowie Darstellungen zu Aufgaben und Einrichtungen des jeweiligen Ressorts; daneben werden jedoch auch Serviceleistungen wie E-Mail-Kontakt und Veranstaltungshinweise angeboten. Alle Ministerien bieten zudem auch einen Zugriff auf Pressemitteilungen, die zum Teil vom Benutzer an die eigene oder fremde E-Mail-Adressen weitergeleitet werden können. Wie im gesamten Internetauftritt ist es auch innerhalb der Pressemitteilungen möglich, über die integrierte Suchmaschine nach Stichworten zu recherchieren.

Pressemitteilungen und Veranstaltungshinweise werden in der Regel täglich aktualisiert, die Informationen zu Aufgaben und Einrichtungen der Ministerien können hingegen zum Teil für mehrere Monate unverändert im Netz stehen, sofern keine Notwendigkeit zur inhaltlichen Überarbeitung besteht. Diese Informationen finden teilweise auch in anderer Form, vornehmlich als Broschüren, Verbreitung. Die Zahl der gedruckten Veröffentlichungen im Bereich des Dienstleistungsangebots und der Öffentlichkeitsarbeit geht jedoch erkennbar zugunsten von Online-Publikationen zurück. Diese Entwicklung wird in naher

¹⁰ Vgl. <http://www.baden-wuerttemberg.de>, <http://www.fm.baden-wuerttemberg.de>, <http://www.im.baden-wuerttemberg.de>, <http://www.jum.baden-wuerttemberg.de>, <http://www.sm.baden-wuerttemberg.de> sowie <http://www.stm.baden-wuerttemberg.de>.

¹¹ Die Titelleiste enthält neben dem Landeswappen und der Bezeichnung des Ministeriums (beim Landesportal der Titel *Baden-Württemberg*) auch stets die Buttons „Home“, „Kontakt“ und „Suchen“; beim Landesportal bieten zusätzlich zwei Schaltflächen die Möglichkeit, eine Version des Internetauftritts in englischer oder französischer Sprache aufzurufen.

¹² URL: <http://www.wir-koennen-alles.de>.

Publikationen zurück. Diese Entwicklung wird in naher Zukunft vermutlich verstärkt zu beobachten sein; das Sozialministerium plant beispielsweise, die Informationsbroschüren, die bisher gedruckt und als PDF-Dateien im Internet erhältlich sind, künftig größtenteils nur noch im Internet zur Verfügung zu stellen.

Zumeist handelt es sich bei den Dokumenten innerhalb der Websites um Textdateien in HTML, es sind jedoch auch Bilder und Grafiken in den Formaten GIF und JPEG vorhanden.¹³ Der Anteil multimedialer Elemente auf den Internetseiten ist verschwindend gering; über das Landesportal etwa kann ein Video mit Ministerpräsident Erwin Teufel eingesehen werden und die Website des Staatsministeriums bietet unter anderem einen virtuellen Rundgang durch das Gebäude der Landesvertretung in Berlin.

Die bei den beschriebenen Websites gewählte Aufteilung in drei Bereiche unterhalb einer Titelleiste findet sich auch bei den Internetseiten des **Ministeriums für Ernährung und Ländlichen Raum (MLR)**¹⁴ und des **Wirtschaftsministeriums**¹⁵.

Auch die MLR-Seiten präsentieren wie das Landesportal die Inhalte zu den vom Benutzer ausgewählten Themen im mittleren Frame. Auf der linken Seite befinden sich allerdings nur Schaltflächen zum Anwählen allgemeiner Themen („Wir über uns“, „Service“ etc.) und ein Bereich zur Volltextsuche; das Menü zur Hauptnavigation befindet sich bei dieser Internetseite im rechten Frame. Bei der Auswahl eines Themas im Hauptmenü erscheint das dazugehörige Untermenü ebenfalls rechts.

Bei der Website des Wirtschaftsministeriums ist das Navigationsmenü als waagrechtlicher Balken unterhalb der Titelleiste platziert. Darunter befinden sich zwei nebeneinander liegende Bereiche, wobei im rechten Frame die vom Benutzer ausgewählten Inhalte dargestellt werden. Auf der linken Seite sind auf der Homepage Buttons zu Aktionen und Initiativen des Ministeriums und des Landes unterge-

¹³ Eine Beschreibung der verschiedenen Formate bietet Raphael Ostermann, *Potentielle Dateiformate zur Langzeitarchivierung von Dokumenten unter Berücksichtigung von Primär- und Metainformationen*, in: Digitale Herausforderung für Archive. Dritte Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 22. und 23. März 1999 im Bundesarchiv Koblenz, hrsg. von Michael Wettengel, Koblenz 1999, S. 25-35.

¹⁴ URL: <http://www.mlr.baden-wuerttemberg.de>.

¹⁵ URL: <http://www.wm.baden-wuerttemberg.de>.

bracht, die wie beim Landesportal durch eine Menüleiste ersetzt werden, sobald ein Thema im Hauptmenü ausgewählt wurde.

Im Gegensatz zu diesen beiden Websites sind bei den Internetauftritten der **Ministerien für Umwelt und Verkehr, für Kultus, Jugend und Sport** sowie **für Wissenschaft, Forschung und Kunst** kaum noch Ähnlichkeiten mit dem Design des Landesportals zu erkennen.¹⁶ Die Seiten des Wissenschaftsministeriums sind zwar mit Ausnahme der völlig anders gestalteten Startseite in drei Bereiche unterteilt, die ähnlich wie beim Internetauftritt des Wirtschaftsministeriums angeordnet sind; die Titelleiste mit Landeswappen und der Bezeichnung des Ministeriums entfällt jedoch beim Blättern durch die Internetseiten. Beim Umwelt- und beim Kultusministerium hat man sich für eine Aufteilung der Seiten in zwei Spalten entschieden. Letztgenannte Website verzichtet weitgehend auf die Darstellung von Untermenüs; Detailinformationen zu den aus dem Hauptmenü ausgewählten Bereichen können häufig über im Text integrierte Hyperlinks aufgerufen werden.

Eine Besonderheit des Online-Angebots des Umweltministeriums ist die an vielen Stellen erfolgende Verknüpfung zur Internetseite der Landesanstalt für Umweltschutz (LfU)¹⁷. Diese Dienststelle im Geschäftsbereich des Ministeriums für Umwelt und Verkehr präsentierte sich früher auf der Website des Ministeriums. Seit Februar 2002 ist die LfU aber nun mit neuer Homepage und erweitertem Angebot im World Wide Web vertreten. Nach wie vor ergeben sich jedoch Überschneidungen bei den Internetangeboten des Ministeriums und der Landesanstalt, etwa bei der Präsentation aktueller Messwerte.

Die Inhalte der fünf zuletzt beschriebenen Internetauftritte entsprechen denen der nach dem Muster des Staatsministeriums gestalteten Websites. In erster Linie werden Informationen über die Organisation, Aufgaben und Einrichtungen der jeweiligen Ministerien vermittelt; zudem wird auch hier auf aktuelle Veranstaltungen und Aktionen hingewiesen sowie die Online-Kommunikation mit den Bürgern und der Zugriff auf Pressemitteilungen ermöglicht. Die Darstellung erfolgt als HTML-Dokumente unter Verwendung von Bildern in GIF und JPEG. Auch PDF-Dateien zum Download werden angeboten.

¹⁶ Vgl. <http://www.uvm.baden-wuerttemberg.de>, <http://www.kultusministerium.baden-wuerttemberg.de> sowie <http://www.mwk.baden-wuerttemberg.de>.

¹⁷ URL: <http://www.lfu.baden-wuerttemberg.de>.

Eine Angleichung des Designs der Websites an das Landesportal strebt keines der fünf zuletzt genannten Ministerien an. Bei künftigen Überarbeitungen der Internetauftritte sollen lediglich Anpassungen in bestimmten Details (z.B. Farbgestaltung) erfolgen. Das Ministerium für Wissenschaft, Forschung und Kunst beabsichtigt allerdings, im Falle einer Neugestaltung der Website des Staatsministeriums den eigenen Internetauftritt dieser Website weitgehend anzugleichen.

Für die Internetpräsentationen aller Ministerien des Landes gilt, dass in der Vergangenheit nur selten grundlegende Überarbeitungen stattgefunden haben. Am häufigsten (dreimal) wurde bisher das Landesportal umgestaltet, das zuletzt Anfang des Jahres 2000 bei der Angleichung an die Website des Staatsministeriums neu strukturiert wurde. Das Sozial- und das Wirtschaftsministerium sowie die Ministerien für Umwelt und Verkehr, für Kultus, Jugend und Sport sowie für Ernährung und Ländlichen Raum haben ihre Internetauftritte bisher einmal neu gestaltet, die anderen fünf Ministerien hingegen noch keine wesentlichen Überarbeitungen ihrer Websites vorgenommen.

Einige der Ministerien planen jedoch bereits eine Neugestaltung ihrer Internetseiten, teilweise noch für das laufende Jahr. Auch im Hinblick auf die Durchsetzung einer EU-Richtlinie zur Schaffung sogenannter *barrierefreier* Internetzugänge¹⁸ ist mit einer Überarbeitung aller Websites der Ministerien in nicht allzu ferner Zukunft zu rechnen.

Die auf den Internetseiten vermittelten Inhalte sind zu einem großen Teil als wichtige historische Quellen einzustufen. Nicht nur, dass viele der Informationen, die Einblicke in die Organisation, die Aufgaben und die Arbeit der Ministerien ermöglichen, lediglich über das Internet verbreitet werden; vielmehr lassen auch die Websites als Ganzes Rückschlüsse darauf zu, wie die Behörden ihre Stellung innerhalb der Landesverwaltung und im gesellschaftlichen Kontext interpretieren und wie sie sich dementsprechend nach außen hin präsentieren.

¹⁸ Damit ist gemeint, behinderten Menschen die Nutzung von Internetangeboten zu ermöglichen. Für sehbehinderte Menschen besteht beispielsweise die Möglichkeit, den Bildschirminhalt über Sprach-Synthesizer akustisch wiedergeben zu lassen; dies gelingt jedoch häufig nicht bei Grafiken, ineinander verschachtelten Frames oder JavaScript-Dateien (vgl. http://www.einfach-fuer-alle.de/navlink_s01_4409.html). Die Behörden der Bundesverwaltung sind durch die *Verordnung zur Schaffung barrierefreier Informationstechnik* dazu verpflichtet, bis spätestens 31.12.2005 ihre Internetauftritte barrierefrei zu gestalten (vgl. http://www.einfach-fuer-alle.de/seite0303_7585.html).

Inwieweit bei einer Archivierung möglicherweise auf einige Elemente der Websites verzichtet werden kann, wird in Kapitel 5 erörtert werden.

3.2 Die Internetseiten des LfD und der Führungsakademie

Der Internetauftritt des **Landesbeauftragten für den Datenschutz (LfD) Baden-Württemberg** wurde Ende 1997 eingerichtet und bisher einmal (im Mai 2000) grundlegend überarbeitet.¹⁹

Der Aufbau der Website ist wesentlich einfacher gestaltet als die Internetseiten der Landesministerien.²⁰ Das Navigationsmenü im linken von zwei nebeneinander angeordneten Frames enthält lediglich sechs Themenbereiche. Bei der Auswahl eines Themas erfolgt dessen Darstellung im rechten Frame zumeist als schlichte Textdarstellung, in die gegebenenfalls weiterführende Hyperlinks integriert sind.

Inhaltlich bieten die Seiten in erster Linie allgemeine Informationen zu den Aufgaben und der Arbeit des LfD. Veranstaltungshinweise oder aktuelle Pressemitteilungen²¹ werden nicht angeboten; eine Aktualisierung einzelner Inhalte muss daher eher selten erfolgen. Über Hyperlinks erfolgt sowohl die Verknüpfung zu den Internetseiten anderer Datenschutzbeauftragter in Bund und Land, als auch die Bereitstellung von Informationsmaterialien, Tätigkeitsberichten sowie von Texten zu Gesetzen und Verordnungen im Zusammenhang mit dem Datenschutz (als HTML-Datei oder als Word-Datei zum Download). Bilder oder Grafiken im JPEG- beziehungsweise GIF-Format werden nur sehr spärlich verwendet.

Nahezu alle Inhalte der Website sind auch in gedruckter Form publiziert, weshalb die Internetpräsentation einer Aneinanderreihung digitalisierter Informationsbrochüren gleicht. Der Wert der Website als historische Quelle muss daher zum gegenwärtigen Zeitpunkt eher als gering eingestuft werden. Eine Archivierung durch das Hauptstaatsarchiv Stuttgart zur Ergänzung des Schriftgutbestandes kommt aus

¹⁹ Folgende Ausführungen basieren, soweit die Informationen nicht dem Internetauftritt entnommen werden konnten, auf einem Anfang März 2003 geführten Gespräch mit der für die Website zuständigen Mitarbeiterin des Landesbeauftragten für den Datenschutz.

²⁰ Vgl. <http://www.baden-wuerttemberg.datenschutz.de>.

²¹ Auf der Website des LfD vom 21. März 2003 stammte die jüngste abrufbare Pressemitteilung vom 9. Dezember 2002.

diesem Grund kaum in Frage.

Die Website der **Führungsakademie Baden-Württemberg** wurde erst Mitte vergangenen Jahres ins Netz gestellt und befindet sich noch immer im Aufbau.²²

Der Internetauftritt ist etwas komplexer gestaltet als jener des LfD.²³ Bei der Auswahl eines der sechs Themen im Hauptmenü auf der Homepage öffnet sich in einem separaten Frame ein Untermenü, das sich gegebenenfalls noch weiter verzweigt. Die jeweils ausgewählten Inhalte werden in einem Frame rechts davon dargestellt. Einer der Bereiche im Hauptmenü beinhaltet Hinweise zu Veranstaltungen und Seminaren der Führungsakademie mit der Möglichkeit, Anmeldeformulare aufzurufen und auszudrucken. Hier finden am häufigsten Aktualisierungen statt. Hinter zwei weiteren Hauptthemen verbergen sich kurz gehaltene Darstellungen zu Aufgaben und Zielen der Führungsakademie, die Bereiche „Informieren“, „Kommunizieren“ und „Aktuelles“ (!) schließlich sind bisher noch ohne Inhalt. Links zu externen Websites sind noch nicht vorhanden, wohl aber künftig vorgesehen.

Viele der bisher ins Netz gestellten Inhalte, deren Darstellung hauptsächlich in HTML erfolgt (Bilder im JPEG-Format), sind auch gedruckt erhältlich, zudem ist der Informationsgehalt der Website insgesamt noch sehr gering. Für die Zukunft ist jedoch nicht auszuschließen, dass die fertiggestellte Internetpräsentation durchaus archivwürdig sein kann. Ob die Website als typisches Beispiel für den Aufbau von Behörden-Internetseiten in regelmäßigen Zeitschnitten archiviert werden sollte, ist fraglich; letztlich würde dadurch nur dokumentiert werden, wie die bisher leeren Menüpunkte nach und nach mit Inhalt gefüllt werden. Außerdem darf nicht übersehen werden, dass der Internetauftritt der Führungsakademie hinsichtlich der hierfür zur Verfügung stehenden Finanz- und Personalmittel nicht vergleichbar ist mit den Internetseiten größerer Behörden.

Es erscheint daher ratsam, eine Bewertungsentscheidung über die Archivwürdigkeit der Website erst dann zu treffen, wenn diese weitgehend vollständig ins Netz gestellt wurde. In die folgenden Überlegungen zur Entwicklung eines Konzepts

²² Folgende Ausführungen basieren, soweit die Informationen nicht dem Internetauftritt entnommen werden konnten, auf einem Anfang März 2003 geführten Gespräch mit der für die Website zuständigen Mitarbeiterin der Führungsakademie Baden-Württemberg.

²³ Vgl. <http://www.fuehrungsakademie.landbw.de>.

zur Archivierung von Internetseiten wurde die Website der Führungsakademie daher nicht einbezogen.

4 Strategien zur Archivierung von Online-Publikationen

4.1 Projekte in Deutschland

Die Deutsche Bibliothek (DDB) hat als *nationalbibliographisches Zentrum* der Bundesrepublik Deutschland die vom Gesetzgeber bestimmte Aufgabe, deutschsprachige und in Deutschland erscheinende Publikationen zu sammeln, zu erschließen und dauerhaft zu sichern. Dieser Auftrag bezieht zwar auch digitale, auf physischen Trägern wie Disketten, CD-ROM und DVD erscheinende Veröffentlichungen mit ein, nicht jedoch Online-Publikationen, die lediglich über Netzwerke verbreitet werden.²⁴

Bis zu einer angestrebten Gesetzesnovellierung zur Ausweitung des Sammlungsauftrags auf den Bereich der Netzpublikationen bemüht sich DDB darum, Online-Publikationen, die ihr auf freiwilliger Basis überlassen werden, zu übernehmen und zu archivieren.

In den Jahren 1998-2000 hat Die Deutsche Bibliothek im Rahmen des Projekts *Networked European Deposit Library - NEDLIB*²⁵ zunächst den dort entwickelten *web harvester* getestet. Es handelt sich hierbei um einen Softwareroboter, der Dateien im Internet aufspüren und ohne technische Unterstützung seitens des Originalservers in den Datenspeicher des Archivs kopieren kann.²⁶

Diese Art des *web harvesting* ermöglicht zwar einerseits eine automatisierte Übernahme und Archivierung von Online-Publikationen, hat andererseits aber den entscheidenden Nachteil, dass die Technik zum heutigen Zeitpunkt noch nicht in der Lage ist, dynamische, auf Content-Management-Systemen²⁷ basierende Websites mit sämtlichen Funktionen in das Archivsystem zu übertragen. Die Deutsche

²⁴ Hier und zu Folgendem vgl. Elisabeth Niggemann / Ute Schwens, *Rahmenvereinbarung Netzpublikationen*, in: *Dialog mit Bibliotheken* 14 (2002), Heft 2, S. 4-6 (hier: S. 4) und <http://www.ddb.de/wir/netzpubl.htm>.

²⁵ URL: <http://www.kb.nl/coop/nedlib>.

²⁶ Vgl. Hans Liegmann, *Selbstbedienung oder Lieferung frei Haus?*, in: *Dialog mit Bibliotheken* 14 (2002), Heft 1, S. 15-18 (hier: S. 17); künftig zit. als: Liegmann.

²⁷ Vgl. hierzu Max Lindlbauer, *Content Management*, in: *Management-Technologien. Konvergenz von Knowledge-, Dokumenten-, Workflow- und Contentmanagement*, hrsg. von Gerhard Versteegen, Berlin u.a. 2002, S. 126-149.

Bibliothek ist daher zu der Überzeugung gelangt, dass derzeit „ein in Qualität und Vollständigkeit befriedigendes Ergebnis durch den Einsatz eines Software-Roboters nicht zu erreichen“ ist.²⁸

In Abkehr vom Selbstbedienungsverfahren wurde daher eine *Arbeitsgemeinschaft Elektronische Depotbibliothek* gebildet. Die an dieser Arbeitsgemeinschaft beteiligten Verleger und Produzenten boten probeweise der Deutschen Bibliothek ihre Online-Publikationen in einem von der Bibliothek vorgegebenen Format zur Archivierung an. Aufbauend auf diesen Erfahrungen wurde schließlich ein Verfahren zur Abgabe von Netzpublikationen durch Verlage und verlegende Stellen unter bestimmten Rahmenbedingungen²⁹ entwickelt.

Seit Oktober 2001 ist das Archivsystem der Deutschen Bibliothek zur Langzeiterhaltung von elektronischen Publikationen mit einer ausgewählten Gruppe von Verlagen³⁰ in Betrieb. Von Interesse für die Archivierung sind in erster Linie:³¹

- Elektronische Publikationen, die durch eine ISSN oder ISBN gekennzeichnet sind und über das Internet verbreitet werden,
- Online-Publikationen, die in keiner anderen Veröffentlichungsform erscheinen,
- Online-Publikationen, deren Sammlung und Archivierung auf keine technische Schwierigkeiten stößt.

Die *Richtlinien für die Sammlung von Netzpublikationen durch die Deutsche Bibliothek*³², die im Hinblick auf die geplante Novellierung des Gesetzes über Die Deutsche Bibliothek erstellt wurden, sehen vor, Websites, die eine Institution oder ein Unternehmen vorstellen (Internetauftritt im Rahmen der Öffentlichkeitsarbeit), als nicht sammelpflichtig einzustufen.

Auch an den Internetseiten staatlicher Behörden besteht derzeit seitens der Deutschen Bibliothek eher untergeordnetes Interesse, da solche in der Regel nicht dem bisherigen Sammlungsprofil und den angestrebten Qualitätsstandards für Online-

²⁸ Liegmann, S. 17 f.

²⁹ Durch § 2 Abs. 1 der Rahmenvereinbarung zwischen der Deutschen Bibliothek und dem Börsenverein des Deutschen Buchhandels wird beispielsweise die unaufgeforderte Ablieferung der Netzpublikationen innerhalb von zwei Wochen nach deren Erscheinen in einem festgelegten Transferformat vorgeschrieben (vgl. http://deposit.ddb.de/netzpub/web_rahmenvereinbarung.htm).

³⁰ Derzeit sind zehn Verlage und verlegende Stellen in die Ablieferung einbezogen (vgl. http://deposit.ddb.de/netzpub/web_liste_ablieferer.htm).

³¹ Zu Folgendem vgl. http://deposit.ddb.de/netzpub/web_rahmenvereinbarung.htm sowie http://deposit.ddb.de/netzpub/web_abgabe_np_verfahren.htm.

³² URL: http://deposit.ddb.de/netzpub/web_sammelrichtlinien.htm.

Publikationen entsprechen. Inwieweit durch die frühestens für das Jahr 2004 zu erwartende Gesetzesnovellierung der Deutschen Bibliothek die Pflicht zur Übernahme von Behörden-Websites zufällt und in welchem Umfang - d.h. für die meist dynamischen Seiten vor allem in welchen Zeitschnitten - eine Archivierung gegebenenfalls vorgeschrieben sein wird, ist derzeit nicht voraus zu sehen.³³

Auf die Online-Publikationen, die bereits auf dem Archivserver der Deutschen Bibliothek gespeichert sind, können berechtigte Nutzer innerhalb der Standorte der Bibliothek in Leipzig, Frankfurt a.M. und Berlin zugreifen. Mit dieser Regelung wollen die bisher an dem Projekt beteiligten Verlage und verlegenden Stellen verhindern, dass die von ihnen häufig kostenpflichtig angebotenen Publikationen frei - etwa über das Internet - verfügbar werden. Nach Möglichkeit sollen jedoch für die Regionalbibliotheken, für welche DDB die Sammlung und Archivierung von Netzpublikationen übernimmt, sogenannte „regionale Fenster“ eingerichtet und somit für diese Bibliotheken der Zugriff auf die entsprechenden Publikationen ermöglicht werden.³⁴

Das **Bundesarchiv**, das bereits auf langjährige Erfahrungen im Umgang mit digitalem Archivgut verweisen kann³⁵, nahm Ende 2001 in Zusammenarbeit mit der Deutschen Bibliothek und dem Presse- und Informationsamt der Bundesregierung ein Projekt zur Archivierung der Website *www.bundesregierung.de* in Angriff.³⁶ Beabsichtigt war, eine Privatfirma mit der Erstellung eines technischen Konzepts zur Archivierung der Internetseite zu beauftragen. Dabei sollte einerseits die komplette statische *Nur-Text-Version* mit allen Änderungen abgespeichert werden, andererseits war aber auch die Speicherung monatlicher *Snapshots* der deutschsprachigen grafischen Version geplant. Die Anforderung an diese *Schnappschüsse* war, dass durch sie die ursprüngliche Struktur der Website wiedergegeben wird

³³ Auskunft von Herrn Hans Liegmann (DDB) vom 25. Februar 2003.

³⁴ Vgl. http://deposit.ddb.de/netzpub/web_rahmenvereinbarung.htm.

³⁵ Vgl. Ulf Rathje, *Technisches Konzept für die Datenarchivierung im Bundesarchiv*, in: Der Archivar 55 (2002), S. 117-120 (künftig zit. als: Rathje, *Technisches Konzept*) sowie ders., *Archivierung von DDR-Daten im Bundesarchiv - ein Rückblick auf zehn Jahre*, in: Mitteilungen Sonderforschungsbereich 580 „Gesellschaftliche Entwicklungen nach dem Systemumbruch. Diskontinuität, Tradition und Strukturbildung“, Jg. 2002, Heft 3 (Heinrich Best / Ronald Gebauer (Hg.), (Dys)funktionale Differenzierung? Rekrutierungsmuster und Karriereverläufe der DDR-Funktionseleiten), S. 15-20 (<http://www.sfb580.uni-jena.de/veroeffentlichungen/heft3.pdf>).

(*Look & Feel*), wobei Browser-Funktionen (wie etwa das Blättern innerhalb der Netzpublikation) erhalten bleiben sollten; auf viele weitere der ursprünglichen Funktionen (Mail, externe Links etc.) wurde hingegen im Hinblick auf den hierzu erforderlichen technischen Aufwand verzichtet.

Leider scheiterte die Durchführung des Projekts bisher an der ungeklärten Finanzierungsfrage. Wann es zur Ausführung gelangen kann, ist angesichts der derzeitigen Lage der öffentlichen Haushalte fraglich.

Das Bundesarchiv plant daher, zunächst einmal die eigene Website zu archivieren, um auf diese Weise Modell-Lösungen für die Archivierung von dynamischen Online-Publikationen zu entwickeln. Beabsichtigt ist, nach der Installation des neuen Internetauftritts des Bundesarchivs im April 2003³⁷ jede Änderung der Seite automatisch zu archivieren. Konkrete Maßnahmen hierfür können jedoch zum jetzigen Zeitpunkt noch nicht in Angriff genommen werden.

Das **Archiv der sozialen Demokratie (AdsD)** unternimmt seit 1999 Anstrengungen im Bereich der Web-Archivierung. Ziel ist es, Internetseiten der SPD, der Friedrich-Ebert-Stiftung und des AdsD dauerhaft zu sichern und für künftige Benutzung zugänglich zu halten.³⁸ Eine Archivierung des Inhalts der Websites als reine *Nur-Text-Versionen* hält das Archiv nicht für ausreichend, da auch Aufbau und Struktur der Seiten überliefert werden sollen. Daher werden die kompletten Internetauftritte mit sämtlichen Funktionen (auch Links) gespiegelt, so dass ein möglichst identisches Abbild erzeugt wird. Anschließend wird die Spiegelung mit einem Index versehen, der eine Suche innerhalb der Seiten ermöglicht. Für die spätere Benutzung wird zusätzlich eine Startseite angelegt, von der aus der Index oder die Suchfunktion gestartet werden kann.

Die (bisher) auf CD-R abgespeicherten Ergebnisse einer derartigen Spiegelung genügen archivischen Ansprüchen in voller Weise und lassen für den Benutzer kaum Wünsche offen. Allerdings ist das bisher angewandte technische Verfahren

³⁶ Folgende Ausführungen basieren auf Informationen, die mir Frau Dr. Andrea Süchting-Hänger (BArch) während meines am Bundesarchiv Koblenz absolvierten Stage (13.01.-7.02.2003) freundlicherweise zur Verfügung stellte.

³⁷ Zur Überarbeitung der Website des Bundesarchivs vgl. Joachim Rausch, *Einführung eines Content-Management-Systems - Neugestaltung des Bundesarchiv Web-Auftritts*, in: Mitteilungen aus dem Bundesarchiv, Jg. 2002, Heft 3, S. 22.

³⁸ Hier und zu Folgendem vgl. Rudolf Schmitz, *Archivierung von Internetseiten / Spiegelungsprojekt im Archiv der sozialen Demokratie (AdsD)*, in: Der Archivar 55 (2002), S. 135 f.

sehr aufwendig und erfordert zu viel „Handarbeit“; so müssen etwa viele Links, die nicht automatisch übertragen werden können, nachträglich manuell eingefügt werden.³⁹

Der hohe Aufwand erklärt wohl auch, warum geplant ist, Spiegelungen des SPD-Internetauftritts lediglich dreimal pro Jahr (Bundesebene) beziehungsweise nur halbjährlich (Landesverbände) vorzunehmen. Als Ergänzung sollen allerdings auch die Websites im Umfeld von Wahlkämpfen auf Landes- und Bundesebene archiviert werden.

4.2 Projekte im Ausland

Als eine der ersten Organisationen weltweit hat sich das 1996 in San Francisco gegründete **Internet Archive**, der Aufgabe angenommen, alle öffentlich zugänglichen Internetseiten möglichst vollständig zu archivieren.⁴⁰

Die Auswahl der Seiten erfolgt durch *web crawler*, d.h. durch Softwareroboter, die wie der bereits erwähnte *NEDLIB-web-harvester* Websites, die zuvor noch nicht archiviert wurden oder in modifizierter Form erscheinen, automatisch aufspüren und als *Snapshot* in das Archivsystem kopieren.

Bisher wurden Daten im Umfang von über 100 Terabytes auf den Archivservern abgelegt; der derzeitige Zuwachs beträgt etwa 12 Terabytes (!) pro Monat.

Die archivierten Seiten werden über das Internet der Öffentlichkeit kostenlos zur Verfügung gestellt. Das von den Machern des Internet Archive gegründete Unternehmen *Alexa Internet* hat hierzu eine als *Wayback Machine* bezeichneten Suchmaschine entwickelt, mit der ältere Versionen aktueller Websites oder aber auch Seiten, die außerhalb des Internet Archive überhaupt nicht mehr im Netz vorhanden sind, aufgerufen werden können.

³⁹ So die Ausführungen von Rudolf Schmitz anlässlich der Präsentation einer gespiegelten SPD-Internetseite im Bundesarchiv Koblenz, Referat B 2 (Überlieferungsbildung aus elektronischen Systemen) am 22. Januar 2003.

⁴⁰ Hier und zu Folgendem vgl. <http://www.archive.org> sowie Jürg Hagmann, «On the dark side of the cyberspace.» *Zur Archivierung des Internets*, in: *Arbido* 14 (1999), Heft 5, S. 14-16 und Hartmut Weber, *Im Internet - aber nicht in Sicherheit. Möglichkeiten und Grenzen der Digitalisierung als Bestandserhaltungsmaßnahme*, in: *Service im Wandel. Bestandssicherung, Elektronische Bibliothek, Veränderungsmanagement*. 27. Arbeits- und Fortbildungstagung der Arbeitsgemeinschaft der Spezialbibliotheken e.V. / Sektion 5 im Deutschen Bibliotheksverband, bearb. von Margit Brauer, Karlsruhe 1999, S. 53-70 (hier: S. 62 f.).

So beeindruckend die Leistung der Betreiber des Internet Archive insbesondere hinsichtlich der archivierten Datenmengen sein mag; die Qualität des Angebots ist aus archivischer Sicht eher unzulänglich. Die bereits angesprochenen technischen Schwierigkeiten des *web harvesting* haben zur Folge, dass viele Websites nur bruchstückhaft archiviert werden. Während beispielsweise die Archivierung von HTML-Seiten offensichtlich nur wenige Probleme bereitet, lassen sich JavaScript-Dateien häufig nicht mit allen Elementen und Funktionen archivieren.

Auch scheint bei der Auswahl der Seiten mehr dem Zufall überlassen zu sein, als dies wohl eigentlich beabsichtigt ist. Von einer (zumindest annähernd) vollständigen Archivierung des Internets kann daher kaum die Rede sein.

Von der Website der Landesregierung Baden-Württemberg www.baden-wuerttemberg.de wurden bisher 56 Versionen durch das Internet Archive archiviert, die älteste vom Dezember 1996, die jüngste derzeit vom November 2002. Einige der laut Index archivierten Versionen sind allerdings überhaupt nicht abrufbar und andere nur sehr unvollständig vorhanden.

Höheren Ansprüchen genügen hingegen etwa die Archivversionen der Websites des Ministeriums des Innern und des Finanzministeriums Baden-Württemberg. Die durch die *Wayback Machine* abrufbaren Seiten dieser Behörden sind jeweils nahezu vollständig archiviert, einschließlich vieler funktionsfähiger Links. Allerdings lässt die Zahl der Versionen etwas zu wünschen übrig. Seit August 2000 (jeweils erste Archivversion) sind in beiden Fällen lediglich sieben Fassungen archiviert worden, vom Jahr 2002 ist sogar jeweils nur eine (im Mai erstellte) Version überliefert.

Einen etwas anderen Weg als das Internet Archive beschreitet die **National Library of Australia**. Diese betreibt mit dem *PANDORA Archive* (**P**reserving and **A**ccessing **N**etworked **D**ocumentary **R**esources of **A**ustralia) ein Archiv ausgewählter australischer Online-Publikationen.⁴¹ Strenge Auswahlkriterien sind die Ursache dafür, dass im Gegensatz zu den zwischen 1996 und 2001 archivierten 10 Milliarden Websites des Internet Archive das PANDORA Archive im gleichen Zeitraum lediglich 1250 Internetseiten übernommen hat. Inzwischen sind allerdings bereits etwas über 3000 Seiten im Umfang von 400 Gigabyte archiviert

worden. Für die Auswahl der Online-Publikationen sind in erster Linie die Inhalte ausschlaggebend, unabhängig von den technischen Anforderungen für die Archivierung. Das PANDORA Archive hat daher neben statischen auch komplexe dynamische Seiten (etwa mit Multimedia-Anwendungen) übernommen, deren Funktionen möglichst vollständig erhalten bleiben sollen. Archiviert werden vor allem:

- Online-Publikationen mit Bezug zu Australien,
- Websites australischer Verfasser,
- Online-Publikationen, denen ein bleibender Wert als historische Quelle zuzuschreiben ist (darunter auch die Veröffentlichungen staatlicher Stellen),
- Online-Publikationen, die in keiner anderen Veröffentlichungsform erscheinen.

Voraussetzung für eine Archivierung durch das PANDORA Archive ist grundsätzlich die Zustimmung des Autors beziehungsweise Verlegers der Online-Publikation. Gegebenenfalls kann mit diesem eine eingeschränkte Nutzung der Archivversion vereinbart werden, etwa die ausschließliche Einsichtnahme innerhalb der staatlichen Bibliotheken oder die Verwendung eines Zugangscodes.

Die Übernahme der Online-Publikationen erfolgt in der Regel durch einen Softwareroboter, der die zur Archivierung bestimmten Websites in den Archivserver kopiert. Dorthin werden auch die erforderlichen Metadaten für die Nutzung und Langzeitarchivierung übertragen. Lediglich sehr komplexe Internetseiten können nicht automatisch archiviert werden; in diesem Fall ist das PANDORA Archive auf die Übermittlung einer Kopie der Seite durch die verlegende Stelle angewiesen. Nach der Überprüfung auf Vollständigkeit und Funktionsfähigkeit wird jede Netzpublikation, für die keine eingeschränkte Nutzung vereinbart ist, schließlich von der National Library der Öffentlichkeit zur Verfügung gestellt.

Die **Norwegische Nationalbibliothek** beschäftigt sich seit einigen Jahren mit der Archivierung von Internetseiten. Im Gegensatz zu den gesetzlichen Bestimmungen in Deutschland sind in Norwegen Online-Publikationen bereits in das Pflichtexemplargesetz einbezogen, was bedeutet, dass der Nationalbibliothek das Recht zur Archivierung jeder norwegischen Website zusteht.⁴²

⁴¹ Hier und zu Folgendem vgl. <http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html> sowie <http://www.nla.gov.au/nla/staffpaper/2002/phillips2.html>.

⁴² Hier und zu Folgendem vgl. Nikola Korb / Berthold Weiß, *The Nordic Web Archive*, in: Dialog mit Bibliotheken 14 (2002), S. 30-32; künftig zit. als: Korb / Weiß.

Zur Entwicklung von Lösungsstrategien für den hierdurch neu entstandenen Aufgabenbereich gründete die Norwegische Nationalbibliothek im November 2000 gemeinsam mit den Nationalbibliotheken der Länder Dänemark, Finnland, Island und Schweden das *Nordic Web Archive*⁴³ mit dem Ziel, „das kulturelle Erbe eines Landes, welches teilweise nur noch als digitale Information über das Internet zugänglich ist, dauerhaft zu archivieren und für die Nachwelt langfristig zur Verfügung zu stellen.“⁴⁴

Die Archivierung der Online-Publikationen erfolgt in einem Depotsystem, das neben der Speicherung der Websites verschiedenster Formate, d.h. von statischen Seiten bis zu Multimedia-Anwendungen, auch die Verwaltung der dazugehörigen Metadaten ermöglicht und schließlich über integrierte Suchmaschinen den Zugang zu den Daten gewährleistet. Zum Auffinden früherer oder späterer Versionen einer Website wurde zusätzlich ein Tool zur Zeitnavigation entwickelt.

Ähnlich wie beim Internet Archive und beim PANDORA Archive werden auch die norwegischen Internetseiten durch Softwareroboter gesammelt und in das Depotsystem überführt. Bei der erstmaligen Speicherung wird ein vollständiger *Snapshot* hergestellt, von bereits existierenden Seiten werden dann gegebenenfalls zusätzlich die Änderungen archiviert.

Für die Zukunft ist geplant, die *Harvesting*-Methoden zu verbessern, um z.B. Änderungen der Seiten kontinuierlich erfassen und archivieren zu können. Zudem soll das Problem „gebrochener“ Links künftig besser als bisher gelöst werden.

⁴³ URL: <http://nwa.nb.no/>.

⁴⁴ Korb / Weiß, S. 30.

5 Konzept für die Archivierung von Websites durch das HSTAS

Wie in Kapitel 3 geschildert wurde, stellen die Websites aller zehn Landesministerien sowie das vom Staatsministerium Baden-Württemberg gepflegte *Landesportal* wichtige historische Quellen dar, die es dauerhaft zu bewahren gilt. Es stellt sich daher die Frage, welche Einrichtung für die Archivierung dieser Internetseiten zuständig ist.

Die Ministerien selbst haben sich größtenteils noch nicht mit dieser Thematik beschäftigt.⁴⁵ In der Regel wurden bisher der Aufbau, die Einrichtung und gegebenenfalls die Änderungen der Internetseiten in Papierform dokumentiert. Einige Ministerien haben Dateien, die in der Vergangenheit von den Websites entfernt worden waren, für einige Wochen aufbewahrt, um eventuell nochmals kurzfristig darauf zurückgreifen zu können. Lediglich bei zwei Landesministerien, beim Ministerium für Ernährung und Ländlichen Raum (MLR) sowie beim Justizministerium, hat man sich bereits mit der Frage der Website-Archivierung auseinandergesetzt. Vom Internetauftritt des MLR werden seit einiger Zeit *Snapshots* auf CD-R gespeichert, ohne dass aber ein Konzept vorliegt, wie mit diesen Archivversionen weiter zu verfahren ist. Eine archivistischen Grundsätzen genügende dauerhafte Aufbewahrung der Websites durch die Ministerien selbst ist für die Zukunft kaum zu erwarten und von diesen vermutlich auch nicht zu leisten. Angesichts der Aufgabe, die Internetseiten zu sichern, zu archivieren, zu erschließen und für die Benutzung zugänglich zu machen, ist eine Eigenarchivierung durch die Ministerien wohl auch kaum anzustreben.

Die Deutsche Bibliothek respektive die für Baden-Württemberg zuständigen Landesbibliotheken in Karlsruhe und Stuttgart haben wie erwähnt bisher keine gesetzliche Verpflichtung zur Übernahme von Online-Publikationen. Die derzeit auf freiwilliger Basis betriebene Sammlung von Netzpublikationen durch DDB bezieht die Internetseiten von Behörden nicht mit ein. Erst die angestrebte Ausweitung des Sammlungsauftrags durch eine frühestens für das Jahr 2004 zu erwartende Gesetzesnovellierung könnte es der Deutschen Bibliothek und den Regional-

⁴⁵ Wie Anm. 9.

bibliotheken zur Pflichtaufgabe machen, auch Behörden-Websites zu übernehmen. Selbst dann jedoch bliebe die Frage, ob angesichts des geringen Stellenwerts, den DDB dieser Art von Netzpublikationen beimisst, die dauerhafte Aufbewahrung der Websites in einem aus archivischer Sicht befriedigenden Umfang erfolgen würde.

Für das Hauptstaatsarchiv Stuttgart, dessen Bestände durch die Archivierung der genannten Websites eine sinnvolle Ergänzung erfahren könnten, erscheint es daher nicht ratsam, sich allein auf die Sammlungstätigkeit der Bibliotheken zu verlassen. Zumindest die derzeitige „Generation“ von Internetauftritten der Landesministerien⁴⁶ wäre ansonsten weitgehend unwiederbringlich verloren, da eine Neugestaltung der derzeitigen Websites noch vor der oben erwähnten Gesetzesnovellierung zu erwarten ist.

Dies schließt natürlich eine Zusammenarbeit mit der Deutschen Bibliothek nicht aus. Vielmehr sollte auf jeden Fall auf das dort vorhandene Know-how zurückgegriffen werden; auch wäre zu überprüfen, inwieweit für die Archivierung der genannten Websites die technische Ausstattung der Deutschen Bibliothek mitbenutzt werden könnte. Die Federführung bei der Auswahl, Erschließung und Archivierung sollte jedoch bei dieser besonderen Form von Websites auf der Seite des Archivs liegen.

Die technischen Bedingungen für die dauerhafte Bewahrung und Erhaltung der Benutzbarkeit der Websites entsprechen im wesentlichen den Voraussetzungen für die Archivierung elektronischer Unterlagen aus Behördenregistraturen. Für die baden-württembergische Archivverwaltung hat Christian Keitel im vergangenen Jahr ein Konzept zur Archivierung elektronischer Unterlagen erstellt⁴⁷; auf technische Details, wie etwa die Frage, ob die Langzeiterhaltung der Dateien⁴⁸ durch Migration (*Refreshing*) oder Emulation gewährleistet werden kann, braucht an dieser Stelle daher nicht eingegangen zu werden. Bei der Archivierung von Onli-

⁴⁶ Immerhin fünf der erwähnten Internetpräsentationen wurden seit der Erstellung der Websites noch nicht grundlegend überarbeitet, so dass man diese noch in der Ursprungsversion sichern und archivieren könnte.

⁴⁷ Vgl. <http://www.lad-bw.de/lad/konzeption.pdf> (Christian Keitel, *Die Archivierung elektronischer Unterlagen in der baden-württembergischen Archivverwaltung. Eine Konzeption*).

⁴⁸ Zur Langzeiterhaltung elektronischer Unterlagen vgl. auch die Ergebnisse des vom Bundesministerium für Bildung und Forschung initiierten Workshops *Langzeitverfügbarkeit digitaler Dokumente* (URL: <http://www.dl-forum.de/foren/langzeitverfuegbarkeit/>).

ne-Publikationen sind jedoch zusätzlich einige Besonderheiten zu beachten⁴⁹:

So stellt sich zunächst die Frage, auf welchem Weg die Websites in das Archivsystem gelangen sollen. Die automatisierte Übernahme durch Softwareroboter direkt aus dem Netz (*web harvesting*) birgt wie erwähnt das Risiko partieller Datenverluste in sich. Aus diesem Grund ist die Übermittlung der Internetseiten mit dazugehörigen Dokumentationsunterlagen und Metadaten durch die Ministerien an das Archiv anzustreben.

Da es sich bei den genannten Internetauftritten um dynamische Netzpublikationen handelt, ist zudem zu erörtern, inwieweit Veränderungen innerhalb der Seiten bei der Archivierung Berücksichtigung finden müssen. Hier ist zunächst daran zu denken, in regelmäßigen Zeitschnitten *Snapshots* der kompletten Websites anzufertigen, um das Aussehen und die Struktur der Seiten zu einem bestimmten Zeitpunkt dokumentieren zu können (*Look & Feel*). Der Konzeption des Bundesarchivs für die Archivierung der Website *www.bundesregierung.de* folgend sollten dabei Browser-Funktionen wie das Blättern innerhalb der Netzpublikation erhalten bleiben, E-Mail-Funktionen und die Verknüpfung zu externen Websites hingegen aufgrund der technischen Anforderungen nicht übernommen werden. Der Wert der Online-Publikation als historische Quelle würde durch diese Einschränkungen nur unwesentlich geschmälert werden. Sollte eine Archivierung der bei allen Internetauftritten integrierten Suchfunktionen technische Schwierigkeiten bereiten, könnte auf diese ebenso verzichtet werden. Für die spätere Benutzung der archivierten Versionen ist es ohnehin sinnvoll, eine Recherchefunktion für die Stichwortsuche im gesamten Website-Bestand, d.h. nicht nur innerhalb einer einzelnen Archivversion, zu ermöglichen.

Die Erstellung von *Snapshots* hätte höchstens monatlich, mindestens jedoch alle drei Monate zu erfolgen sowie unmittelbar nach jeder grundlegenden Überarbeitung der Internetseiten. Außerdem sollten fortlaufend alle inhaltlichen Änderungen (Aktualisierungen) der Seiten von den Ministerien als *Nur-Text-Versionen* gespeichert und mit entsprechenden Angaben über die Platzierung der Texte innerhalb der Websites an das Archiv übergeben werden.

⁴⁹ Vgl. hierzu auch Stephen Harries, *Capturing and managing electronic records from websites and Intranets in the government environment*, in: INSAR, Supplement IV (2000): Proceedings of the DLM-Forum on electronic records. European citizens and electronic information: the memory of the Information Society, S. 72-76 (Zusammenfassung in deutscher Sprache S. 76 f.).

Die Speicherung der zu archivierenden Websites sollte möglichst in standardisierten, flachen Formaten erfolgen, ohne jedoch bei den *Snapshots* unerwünschte Informationsverluste zu verursachen. Die *Nur-Text-Versionen* könnten als ASCII-CSV-Dateien gespeichert werden, die bei der Archivierung (v.a. hinsichtlich einer verlustfreien Migration) verhältnismäßig geringe Probleme bereiten. Bei Neugestaltungen der Websites sollte die archivierende Stelle künftig bereits im Vorfeld informiert werden, um Einfluss auf die Gestaltung im Hinblick auf die Langzeiterhaltung (z.B. Vermeidung von Multimedia-Anwendungen) nehmen zu können. Ob die Langzeitarchivierung der Websites innerhalb eines komplexen elektronischen Depotsystems, wie es die Königliche Bibliothek der Niederlande eingerichtet hat⁵⁰, erfolgen sollte oder auf Wechseldatenträgern wie CD-R und DAT, wie es das Bundesarchiv praktiziert⁵¹, ist nicht zuletzt eine Frage der zur Verfügung stehenden Finanzmittel. Wie erwähnt sollte hinsichtlich der Archivierung eine Kooperation mit der Deutschen Bibliothek angestrebt werden, die ihrerseits bereits Online-Publikationen archiviert hat. Im Falle einer Eigenarchivierung durch das HSTAS wäre angesichts der (bisher) überschaubaren Menge an archivwürdigen Websites die Verwendung von Wechseldatenträgern ins Auge zu fassen. Dies wäre mit einem verhältnismäßig geringen Kostenaufwand verbunden und würde auch hinsichtlich der Unterbringung im Magazin keinen größeren Aufwand verursachen.⁵²

Für die Benutzung schließlich könnten entweder Kopien in benutzerfreundlichen Formaten auf Wechseldatenträgern gespeichert, oder aber diese Kopien online zur Verfügung gestellt werden. Letztere Möglichkeit bietet sich bei dieser besonderen Form digitaler Unterlagen an, da im Gegensatz zu dem aus den Behördenregistra-turen stammenden Archivgut die Websites ursprünglich ohne Einschränkung öffentlich zugänglich waren und daher in der Regel keine Nutzungsbeschränkungen vorzusehen sind.

Der Workflow für die Übernahme einer als archivwürdig bewerteten Website durch das HSTAS könnte folgendermaßen aussehen:

⁵⁰ Vgl. <http://www.kb.nl/e-depot>.

⁵¹ Vgl. Rathje, *Technisches Konzept*, S. 119 f.

⁵² Vgl. ebd.

1. Übergabe der Dateien (möglichst in archivtauglichen Formaten) durch die Behörde an das HSTAS auf elektronischem Weg oder auf Wechseldatenträgern:
 - In zuvor festgelegten regelmäßigen Abständen (sowie unmittelbar nach jeder grundlegenden Überarbeitung der Website) Ablieferung von *Snapshots* (ohne E-Mail-Funktionen und Links zu externen Websites, gegebenenfalls auch ohne Suchfunktionen) mit allen zugehörigen Metadaten und Dokumentationen
 - Monatliche Ablieferung aller seit der letzten Übergabe auf der Website vorgenommenen inhaltlichen Änderungen als *Nur-Text-Versionen* mit Dokumentationen zum Zeitpunkt der Änderung, zur Plazierung der Texte innerhalb der Website etc.
2. Überprüfung der Dateien, Metadaten und Dokumentationen auf Vollständigkeit und Funktionsfähigkeit.
3. Sicherung und Archivierung der übernommenen Dateien (unter Berücksichtigung der Migrationsstrategie zur Langzeiterhaltung):
 - Migration in für die Archivierung geeignete Formate (sofern nicht durch die abgebende Stelle bereits erfolgt), d.h. Erstellung einer Archivversion
 - Metadaten-Datenbank für die Langzeitarchivierung und Nutzung anlegen beziehungsweise ergänzen
 - Übertragung der Daten auf ein Depotsystem oder auf für archivische Zwecke geeignete Wechseldatenträger
 - Erstellung einer Sicherungskopie auf einem anderen Datenträger als die (erste) Archivversion
 - Gegebenenfalls Magazinierung der Wechseldatenträger
 - Regelmäßiges *Refreshing* der Dateien
4. Erschließung und Benutzung:
 - Dateien ordnen
 - Verknüpfung zwischen den Dateien und den Metadaten herstellen
 - Virtuelle Verknüpfung zwischen den Dateien und den (analogen) Dokumentationen herstellen
 - Erstellung dynamischer, d.h. erweiterbarer Findmittel

- Erstellung einer Kopie der Archivversion in einem für die Benutzung geeigneten Format (Nutzungskopie)
- Bereitstellung der Nutzungskopie auf Wechseldatenträgern oder online

Nach einer ersten Phase der Erprobung dieses Konzepts mit der Übernahme von Websites der Landesministerien durch das HSTAS wäre schließlich zu überprüfen, inwieweit eine Erweiterung des Projekts hinsichtlich der Archivierung des Internetangebots der Staatlichen Archivverwaltung Baden-Württemberg und der Internetauftritte von Behörden im Zuständigkeitsbereich anderer Staatsarchive in Frage kommen könnte. Auch eine Einbeziehung der *Intranet*seiten der Behörden wäre in Betracht zu ziehen.

Darüber hinaus ist daran zu denken, auch Websites nichtamtlicher Stellen (Verbände, Vereine, Wirtschaftsunternehmen etc.) einer archivischen Bewertung zu unterziehen. Sofern eine Übernahme derartiger Online-Publikationen im Bereich der Ergänzungsdokumentation, insbesondere der nichtstaatlichen Überlieferungsbildung, von Interesse sein sollte, wäre jedoch der oben dargestellte Workflow wohl um einige Punkte zu erweitern. Vor allem wären im Vorfeld der Übernahme urheberrechtliche Fragen zu erörtern, um spätere Unstimmigkeiten zwischen den Produzenten der Internetseiten und dem Archiv zu vermeiden.⁵³

Koordiniert werden sollte die Website-Archivierung schließlich auch bei einer Ausweitung des Aufgabengebiets durch eine zentrale Stelle, unabhängig davon, durch welches Staatsarchiv die jeweilige Bewertung vorzunehmen ist. Die Benutzung jedoch könnte dezentral erfolgen und/oder vorzugsweise online im Rahmen des Internetangebots der Landesarchivdirektion Baden-Württemberg und der Staatsarchive des Landes.

⁵³ Vgl. hierzu grundlegend Alexander Klett, *Urheberrecht im Internet aus deutscher und amerikanischer Sicht*, Baden-Baden 1998 (Schriftenreihe des Archivs für Urheber-, Film-, Funk- und Theaterrecht 149).

6 Literatur

Literatur zum Thema *Archivierung des Internets* ist bisher leider nur sehr spärlich zu finden. Zumeist werden entsprechende Projekte auf den Websites der beteiligten Institutionen vorgestellt; gedruckte Veröffentlichungen sind nur in geringem Umfang vorhanden.

Die Bemühungen der *Deutschen Bibliothek* zur Archivierung von Online-Publikationen sind verhältnismäßig gut dokumentiert. **Hans Liegmann** befasst sich in einem Aufsatz (*Selbstbedienung oder Lieferung frei Haus?*, in: *Dialog mit Bibliotheken* 14 (2002), Heft 1, S. 15-18) mit diesem Projekt und beleuchtet kritisch die Methode des *web harvesting*. Ergänzt werden seine Ausführungen durch einen Beitrag von **Elisabeth Niggemann und Ute Schwens** (*Rahmenvereinbarung Netzpublikationen*, in: *Dialog mit Bibliotheken* 14 (2002), Heft 2, S. 4-6) über das derzeit angewandte Verfahren zur (freiwilligen) Abgabe von Netzpublikationen durch Verlage und verlegende Stellen an DDB.

Das wohl ambitionierteste Projekt im Bereich der Internet-Archivierung, das in San Francisco angesiedelten *Internet Archive*, beschreibt **Jürg Haggmann** (*«On the dark side of the cyberspace.» Zur Archivierung des Internets*, in: *Arbido* 14 (1999), Heft 5, S. 14-16) kritisch aus der Sicht eines Archivars. Vor allem bemängelt er den recht sorglosen Umgang der Betreiber des Internet Archive mit Urheberrechten und die fehlende Erschließung der archivierten Websites.

Mit der Problematik der Archivierung von Online-Publikationen staatlicher Stellen befasst sich der Brite **Stephen Harries** (*Capturing and managing electronic records from websites and Intranets in the government environment*, in: *INSAR, Supplement IV* (2000): *Proceedings of the DLM-Forum on electronic records. European citizens and electronic information: the memory of the Information Society*, S. 72-76), der im Public Record Office mit der Übernahme elektronischer Unterlagen aus der staatlichen Verwaltung Erfahrungen sammeln konnte. Er beschreibt die Besonderheiten dieser Form digitaler Unterlagen und die Schwierigkeiten bei der Archivierung dynamischer Dokumente.

Ein für die staatliche Archivverwaltung Baden-Württemberg taugliches Konzept zur Archivierung von e-records stellt **Christian Keitel** (*Die Archivierung elektronischer Unterlagen in der baden-württembergischen Archivverwaltung. Eine Kon-*

zeption) auf den Internetseiten der Landesarchivdirektion vor (URL: <http://www.lad-bw.de/lad/konzeption.pdf>). Websites sind darin allerdings unberücksichtigt. Für die Entwicklung eines Konzepts zur Übernahme von Internetseiten amtlicher Stellen durch baden-württembergische Staatsarchive sind jedoch seine Überlegungen unbedingt mit einzubeziehen.

Die technischen Voraussetzungen zur Langzeiterhaltung elektronischer Unterlagen beschreibt schließlich **Ulf Rathje** (*Technisches Konzept für die Datenarchivierung im Bundesarchiv*, in: *Der Archivar* 55 (2002), S. 117-120). Er schildert die Erfahrungen, die am Bundesarchiv mit der Archivierung von Computerdateien aus der ehemaligen DDR gemacht wurden, und befasst sich mit der Frage, in welchen Formaten und auf welchen Datenträgern Dateien gesichert werden müssen, um dauerhaft aufbewahrt werden zu können. Eine Anwendung dieses technischen Konzepts auf die Archivierung von Websites ist jedoch nur bedingt möglich, da dynamische Netzpublikationen anders aufgebaut sind als die am Bundesarchiv archivierten Dateien, die überwiegend datenbankähnliche Strukturen aufweisen.

7 Anhang: Abkürzungsverzeichnis

AdsD	Archiv der sozialen Demokratie
BArch	Bundesarchiv
CD-R	Compact Disc Recordable
CD-ROM	Compact Disc Read-Only Memory
CSV	Comma separated values
DAT	Digital Audio Tape
DDB	Die Deutsche Bibliothek
DDR	Deutsche Demokratische Republik
DVD	Digital Video Disc / Digital Versatile Disc
EU	Europäische Union
GIF	Graphics Interchange Format
HSTAS	Hauptstaatsarchiv Stuttgart
HTML	Hypertext Markup Language
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
JPEG	Joint Photographic Experts Group
LfD	Landesbeauftragter für den Datenschutz
LfU	Landesanstalt für Umweltschutz
MLR	Ministerium für Ernährung und Ländlichen Raum
NEDLIB	Networked European Deposit Library
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia
PDF	Portable Document Format
SPD	Sozialdemokratische Partei Deutschlands
URL	Uniform Resource Locator