

Webarchivierung *Öffentlicher Workshop*

Nestor und der AWV-Arbeitskreis „Dokumentation und Archivierung von Webpräsenzen“ hatten gemeinsam zu einem öffentlichen Workshop eingeladen, um über den derzeitigen Stand der Webarchivierung zu berichten und Probleme und Herausforderungen, die mit der Erfassung, Erschließung, Bereitstellung und Langzeitsicherung der noch jungen Quellengattung verbunden sind, zu diskutieren. 140 Teilnehmer waren der Einladung zu der ganztägigen Veranstaltung gefolgt, die am 20.3.2012 in den Räumen der Deutschen Nationalbibliothek stattfand. Nachdem Reinhard Altenhöner im Namen der Deutschen Nationalbibliothek die Teilnehmer begrüßt hatte, gaben Sabine Schrimpf (Deutsche Nationalbibliothek) und Rudolf Schmitz (Friedrich-Ebert-Stiftung) einen Einblick in die Arbeit sowohl von Nestor und den nestor-workshops zur Webarchivierung als auch dem AWV und dem Arbeitskreis 6.2.

Die Arbeitsgemeinschaft für wirtschaftliche Verwaltung, die als Forum mit mittlerweile über 500 ehrenamtlichen Experten den Dialog zwischen Wirtschaft, Verwaltung und Dienstleistungssektor auf vielfältige Weise organisiert, bietet seit 2007 mit dem Arbeitskreis 6.2 den organisatorischen Rahmen auch für Archive und Bibliotheken, die sich mit Fragen der Webarchivierung befassen.

Zunächst als informelle Weiterführung des DFG-Projekts (2004 – 2006) der Archive der politischen Stiftungen gedacht, erweiterte sich der Teilnehmerkreis in kurzer Zeit erheblich. Mit der Zahl der vertretenen Archivsparten haben sich auch die Anforderungen an die Webarchivierung entsprechend ausdifferenziert.

Der Kooperationsverbund nestor wurde im Rahmen eines vom Bundesministerium für Bildung und Forschung geförderten Projektes zur digitalen Langzeitarchivierung (2003-2009) begonnen und wird seit Juli 2009 von den ehemaligen Projektpartnern gemeinsam mit weiteren Einrichtungen selbstständig weitergeführt. Von Beginn an hat nestor auf seiner Informationsplattform auch den verschiedenen Projekten zur Webarchivierung vielfältige Möglichkeiten der Darstellung eingeräumt, die seit 2011 mit den workshops zur Webarchivierung eine konkrete Fortsetzung und Ergänzung für einen intensiveren Erfahrungsaustausch erhalten haben.



Sabine Schrimpf (Deutsche Nationalbibliothek)
und Rudolf Schmitz (Friedrich-Ebert-Stiftung)..
Foto: Jürgen Klocke, 2012

Beide Veranstalter betonten, dass die Arbeitskreise allen offenstehen, die bereits mit der Webarchivierung befasst sind bzw. sich vor die Notwendigkeit gestellt sehen, damit zu beginnen.

Vor dem Hintergrund konkreter Archivierungsaufträge und technischer Bedingungen wurden in einer ersten Runde exemplarisch von Vertretern verschiedener Archivsparten und dem Landesbibliothekszentrum Rheinland-Pfalz konzeptionellen Lösungen vorgestellt sowie der jeweilige technische und personelle Aufwand für die Webarchivierung beschrieben¹. (Lars Jendral, Landesbibliothekszentrum Rheinland-Pfalz für die Bibliotheken; Johannes Renz, Landesarchiv Baden-Württemberg für die öffentlichen Archive; Ute Schiedermeier, Siemens für die Wirtschaftsarchive; Dr. Renate Höpfinger, Hanns-Seidel-Stiftung für die Parlaments- und Parteiarchive und Robert Fischer, Südwestrundfunk für die Produktionsarchive.)

Die im weiteren Verlauf des Workshops vorgestellten und analysierten Arbeitsschritte der Webarchivierung wurden ergänzt durch die Darstellung der mit der Erfassung und Bereitstellung verbundenen rechtlichen Fragen sowie einen Überblick über den Stand der Diskussion zur Langzeitsicherung archivierter Websites.

Mit den grundlegenden Problemen der Erfassung sowohl der Primärdaten als auch der Sekundärdaten befassten sich gleich zwei Vorträge.



Rudolf Schmitz (Friedrich-Ebert-Stiftung)
Foto: Jürgen Klocke, 2012

Rudolf Schmitz (FES, AdsD) gab einen Überblick über die spezifischen Aufgabenstellungen der unterschiedlichen Erfassungsmethoden und die Besonderheiten ihrer Anwendung. Ausführlicher wurde neben der darstellungsorientierten Spiegelung die vorgangsorientierte Dokumentation durch die sogenannte transaktionale Erfassung dargestellt. Der große Vorteil der Spiegelung als Erfassungsmethode liegt darin, dass sie vollständig clientseitig durchgeführt werden kann und außer dem Einverständnis keine aktive Mithilfe des Anbieters erfordert. Dies und die Darstellungsorientierung unterscheidet sie von allen anderen Erfassungsmethoden. Im Weiteren wurden unterschiedliche Strategien der Spiegelung geschildert, mit denen verhindert werden soll, dass sich Seiten in einem Archivobjekt befinden, die nie gleichzeitig im Web standen.

¹ Die Präsentationen sind im Netz abrufbar unter:
<http://www.langzeitarchivierung.de/Subsites/nestor/DE/Veranstaltungen/vergangene/Webarchivierung2012.html>



Sebastian Vesper (Friedrich-Ebert-Stiftung) im Auditorium
Foto: Jürgen Klocke, 2012

Die Darstellung und Auswertung von serverseitigen Informationen, die beim Aufruf einer Webseite zusätzlich zu den Primärdaten übertragen werden, war Gegenstand einer Analyse der Server-Client-Kommunikation durch Sebastian Vesper (FES, Bibliothek). Die Archivierung auch der übertragenen Header-Informationen mit den darin enthaltenen Angaben zum letzten Änderungsdatum, zur Größe und zum Dateityp der Primärdateien ist unentbehrlich für den Nachweis der Authentizität der in den Archivobjekten aufbewahrten Daten. Die Visualisierung dieses Datentransfers, die von verschiedenen Browsertypen unterstützt wird, kann auch zur Überprüfung der Vollständigkeit bereits gespiegelter Websites eingesetzt werden.

Barbara Signori (Schweizerische Nationalbibliothek) stellte die verschiedenen Formen der Erschließung im Webarchiv Schweiz vor, das in Zusammenarbeit vor allem mit den Kantonsbibliotheken aufgebaut wird. Die Metadaten sowohl aus den Angaben der Partnerinstitutionen als auch aus dem Ingest liegen dem Nachweis im Online-Katalog Helveticat zugrunde, während e-Helveticum dem Benutzer neben einem direkten Zugriff und einer Volltextsuche noch zahlreiche weitere Such- und Filtermöglichkeiten bietet.

In welchen engen rechtlichen Grenzen sich sowohl die Erfassung als auch die Aufbe-

wahrung und Bereitstellung archivierter Websites außerhalb lizenzrechtlicher Vereinbarungen bewegt, machte der Vortrag von Dr. Eric W. Steinhauer (FernUni Hagen) deutlich. Auch auf die Notwendigkeit der Beachtung zahlreicher Aspekte des Persönlichkeitsrechts wurde nachdrücklich hingewiesen. Bei aller Unterschiedlichkeit der Sammlungsaufträge und der rechtlichen Voraussetzungen, unter denen Archive und Bibliotheken die Webarchivierung betreiben, bleibt es eine gemeinsame und wohl eher langfristige Aufgabe, auf eine entsprechende Novellierung des Urheberrechts hinzuwirken.

Tobias Steinke (Deutsche Nationalbibliothek) präsentierte eine Übersicht über den Stand der Diskussion der „Preservation Working Group“ des IIPC, nachdem er mit der Verlinkung der einzelnen Webseiten auf der einen Seite und den zur Darstellung benötigten Plug-ins auf der anderen, zwei Elemente benannt hatte, die einerseits der Migration und andererseits der Emulation als Verfahren der Langzeitsicherung entgegenstehen. Erschwert wird die Entwicklung von Strategien zur Langzeitsicherung auch durch die sich verändernden Formate und Versionen von Dateitypen, die im Web Verwendung finden. Ausgiebig wurden daher Verfahren zur Formatidentifikation, Validierung und Metadatengenerierung vorgestellt und diskutiert.

Im vorletzten Beitrag stellte Angela Ullmann (Parlamentsarchiv des Deutschen Bundestags) das Webarchiv des Deutschen Bundestages vor, das bereits online zugänglich ist. Das Archiv bietet neben komfortablen Zugangsmöglichkeiten, die noch ergänzt werden sollen, umfangreiche Möglichkeiten der Recherche. Die Links zu den einzelnen Seiten im Archiv können aktiv vom Benutzer abgerufen werden oder zur Beantwortung von Anfragen separat verschickt werden. Änderungen, die aufgrund rechtlicher Vorgaben oder Einsprüche

Betroffener in der Benutzungsversion der Archivobjekte vorgenommen wurden, werden in einer für den Benutzer transparenten Weise dokumentiert und angezeigt.

In seinem Vortrag „Webarchive - Eine neue Quelle für die Forschung“ warf Prof. Dr. Hanns Jürgen Küsters (Konrad-Adenauer-Stiftung, ACDP) aus der Sicht des Historikers einige quellenkritische Fragen auf und skizzierte die Anforderungen in Bezug auf eine intellektuelle Erschließung digitaler Archivobjekte, die von ihm in den Mittelpunkt gerückt wurde.

Resümierend wurde in der abschließenden Diskussion noch einmal danach gefragt, welche neuen Standards die Webarchivierung erfüllen muss und welche Standards aus dem Bereich konventioneller Archivierung übernommen, verändert oder ergänzt werden müssen.

Die Beiträge des Workshops sollen noch in diesem Jahr in einer Sondernummer der „AWV-Informationen“ veröffentlicht werden.

Rudolf Schmitz
Bonn, den 2.5.2012