

Die Aufbewahrung von Website-Inhalten

In diesem Beitrag:

- Webarchivierungspflicht nur eine Frage der Zeit
- Welche Webseiten archiviert werden sollten
- Langzeitarchivierung nach ISO-Standard



UBS archiviert Webseiten mit PDF/A

Es gibt keinen Grund, den Inhalt einer Website anders zu behandeln als andere Inhalte. Die Webarchivierungspflicht scheint daher nur eine Frage der Zeit. Die UBS archiviert die Inhalte ihrer Homepage bereits heute.

Infos zum Autor

**Manfred Weise**

Der Autor ist IT-Fachjournalist und Sozialwissenschaftler

Die Archivierung von elektronischen Geschäftsakten ist in den meisten Unternehmen längst «business as usual». Doch wie sieht das mit den Webseiten aus? Das Konzernarchiv für Langzeitaufbewahrung der Grossbank UBS, für die Archivierung von Dokumenten aus dem Führungsprozess verantwortlich, kam in dieser Frage zum Schluss: Die UBS will noch in 5 bis 20 Jahren nachweisen können, was sie heute über ihre Home Page kommuniziert - trotz Disclaimer (Haftungsbeschränkung) auf der Website.

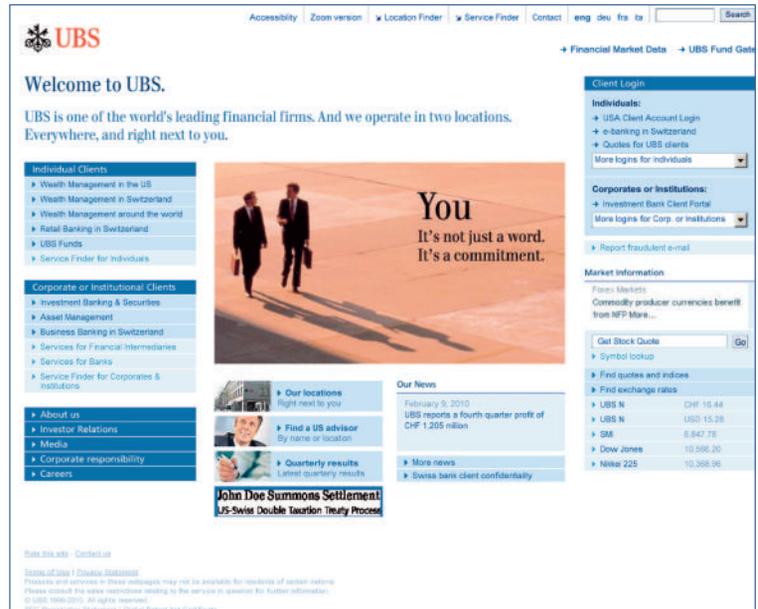
Nicht ganz einhellig ist hierzulande, ob Bestimmungen etwa in der Geschäftsbücherverordnung bereits für Webseiten gelten. Daniel Spichty, UBS-Projekt-

leiter für die Webarchivierung: «Es gibt in der Schweiz zur Zeit keine expliziten rechtlichen Bestimmungen bezüglich einer Aufbewahrung von Webseiten. Aus meiner Sicht wird sich die Rechtsprechung aber ändern. In der Europäischen Union und USA ist diese Tendenz schon sichtbar. Es herrscht dort die juristische Meinung, dass es keinen Grund gibt, den Inhalt einer Website anders zu behandeln als andere Inhalte».

Auf jeden Fall bringt die Archivierung für Spichty nicht nur eine grössere Sicherheit für den Nachweis und die Überprüfung der Kommunikationsinhalte, sondern ist auch für die Bewahrung der Firmengeschichte von grosser Bedeutung. Im Laufe eines



Daniel Spychti, bei der UBS Projektleiter für die Webarchivierung



Webarchivierungsprojektes bei der UBS wurde klar, dass nicht alle Inhalte archivwürdig sind (es macht etwa keinen Sinn Werbetexter im Archiv abzulegen) oder technisch archivfähig gemacht werden können. Aus diesem Grunde wurden Webseiten definiert, die archiviert werden. Dazu gehören auch alle Link-Funktionen.

Die Langzeitarchivierung von Webseiten

Die zu archivierenden Webseiten werden eins zu eins in PDF/A konvertiert. Es entsteht dabei eine rollende Seite im PDF/A-Format, wobei die Paginierung gesteuert werden kann. Hinter dem Volltextsuche unterstützten PDF/A (A steht für Archivierung) versteckt sich der ISO-Standard (International Standards Organisation) für die Langzeitarchivierung –

die Norm stellt sicher, dass sich Dokumente genauso in zehn Jahren noch lesen lassen. «In den PDF/A Dokumenten ist auch die URL hinter einem Link extrahierbar. Damit kann man im Archiv zu weiteren Seiten navigieren, wenn wir die Links verfolgen. Das ist der Grund, warum wir kein reines Bildformat wie Tiff oder JPEG verwenden konnten», so Spichty. Ebenfalls nicht überzeugen konnte die Archivierung von HTML, wie auch Techniken wie Crawling (ein Webcrawler ruft die Inhalte einer Website wie ein Nutzer ab und stellt die Seiten in ein Archiv) oder Harvesting (es werden Webteile mit hohen Zugriffsraten archiviert) verworfen wurden.

Bei der Software-Suche wurde die UBS nach einer längeren Evaluation vor der Haustür fündig. Die Wahl fiel auf den 3-Heights Document Converter Service des Unternehmens «PDF Tools», das im Zürcherischen Winkel beheimatet ist. «Da wir mit dem Projekt sowohl innerhalb als auch ausserhalb der Bank Neuland betreten haben, war es uns wichtig, einen verlässlichen Partner ins Boot zu nehmen. Zudem stellte PDF Tools auch den technologischen Support jederzeit sicher,» sagt Spichty.

Die Suche der identifizierten Inhalte läuft über die URL und das Datum. Mit jeder Änderung erkennt das System automatisch, ob eine neue Archivversion erstellt werden muss oder nicht. Der skalierbare 3-Heights Document Converter Service unterstützt weiter OCR, Verschlüsselung, digitale Signatur und Schnittstellen wie C, Java, .NET, COM. Mit der Software lassen sich Dateiformate nach PDF und PDF/A wie auch nach Tiff konvertieren.

Mit einer Volltext-Suchfunktion wird das Webseiten-Archiv der UBS analog zu den bereits bestehenden Archivgütern geführt.

Info: www.pdf-tools.com

PDF Tools – eine Schweizer IT-Erfolgsgeschichte

Das Zürcher Unternehmen PDF Tools bietet seit mehreren Jahren eine breite Palette an client- und serverbasierten Tools für die Ablage von Dokumenten in PDF und PDF/A. Sie lassen sich von Entwicklern, Integratoren und IT-Abteilungen leicht in andere Applikationen integrieren. Zum Teil wird die Software über OEM-Partner wie Finjan und Docubase vertrieben. Spezialität des Unternehmens ist der ISO-Archivstandard PDF/A; sämtliche Dokumentenarten sind nach PDF/A konvertierbar. Firmeninhaber Hans Bärffuss ist Mitglied der Technical Workinggroup 171 der ISO (International Standards Organisation), das über PDF/A bestimmt. Zu den Kunden gehören Firmen wie IBM, Swisscom und Storz Medical, aber auch Krankenkassen, Versicherungen und öffentliche Verwaltungen. Der Grossteil der Kunden ist international.

