

Repositorien - Digitalmüll der Zukunft oder Zukunft digitaler Bibliotheken?

G. Schneider

Die Vorteile digitaler Dokumente liegen auf der Hand: man kann die Daten, die auf einem Server liegen, quasi mit Lichtgeschwindigkeit über das Internet an den eigenen Arbeitsplatz kopieren. Die Präzision dieser umständlich klingenden Aussage ist beabsichtigt, insbesondere im Licht der aktuellen Tagespolitik mit ihren Patentrezepten für das Internet.

Da in der Vergangenheit Dokumente auf schwer beweglichem Papier erschienen sind, wird in vielen Projekten die Digitalisierung der vorhandenen Schätze - und da gibt es viele - vorangetrieben. Ein bekanntes und erfolgreiches Vorhaben ist das von der DFG geförderte VD16/VD17. Ein Nutzer steht aber viel zu oft vor dem Problem, herausfinden zu müssen, ob ein gesuchtes Dokument bereits digitalisiert ist und wo es gefunden werden kann.

Der Ruf nach der zentral ordnenden Hand erschallt zwar immer wieder - auch in Form einer Projektförderung -, ist jedoch aufgrund der Natur des Internet und der menschlichen Sozialisierung (insbesondere in einem föderalen System) zum Scheitern verurteilt.

Immerhin sind Repositorien entstanden, um dem lokalen Vergessen entgegenzuwirken und die Auffindbarkeit zu steigern. Repositorien sind lediglich geordnete Datensammlungen mit einer mehr oder weniger definierten Verantwortlichkeit. Eine Datensammlung, die von einer Bibliothek betrieben wird, genießt aufgrund der Vorgeschichte und des Ansehens dieser Einrichtung einen Vertrauensvorschuss in Bezug auf die langfristige Verfügbarkeit, der im Einzelfall aber nicht begründet sein muss und den Streichungen der Finanzmittel durch die Geldgeber hilflos ausgesetzt sein wird.

Das Urheberrecht ist überraschenderweise eine große Bedrohung für Repositorien. Viele Datensammlungen sind heute unvollständig, weil die Rechte an jüngeren Digitalisaten nicht geklärt werden können - oder schlichtweg nicht übertragen werden. Öffentliche Institutionen verfügen nicht über die Möglichkeit großer Firmen wie Google, sich im Streitfall das Recht einfach zu kaufen - bzw. den begangenen Rechtsbruch ohne Folgen zu heilen. Ein jüngst im Spiegel (<http://www.spiegel.de/spiegel/0,1518,709761,00.html>) erschienener Artikel beschreibt überraschend negative Folgen des Urheberrechts in der Geschichte.

Die Zentralisierung auf wenige Betreiber von Repositorien - also Besitzern von Daten und damit Information - ist per se gefährlich. Nicht nur, weil im Orwellschen Sinne störende Dokumente einfach ohne Spuren ausgetauscht werden können - wie oft werden Versuche der Geschichtsfälschungen durch Funde vergessener / nicht vernichteter Dokumente erkannt? Diese Wenigen könnten auf die Idee verfallen, den Zugriff auf ihre Datensammlungen zu reglementieren und einzuschränken. Der Ausschluss irakischer Wissenschaftler von der Nutzung US-amerikanischer Wissensdatenbanken ist ein Beispiel hierfür, wenngleich die Ursache in diesem Fall für uns verständlich erscheint. Ohne belastbare langfristige Verträge, sowohl national wie international, und entsprechender gesetzlicher Grundlagen sind von privater Hand oder ausländischen Institutionen betriebene Repositorien nicht ohne Gefahr für die Gesellschaft.

Dagegen sehen die technischen Herausforderungen nahezu beherrschbar aus. Fast allen Nutzern ist das technische Problem der Trägermedien vertraut: Daten, die einst mit Liebe auf Disketten geschrieben wurden, sind heute nicht mehr lesbar - weil es keine Laufwerke

mehr gibt. Die CD-ROM, einst als dauerhaftes Speichermedium gepriesen, weist eine beschränkte technische Lebensdauer von 1-50 Jahren auf und wird dabei trotzdem vermutlich die Verfügbarkeitsdauer der Lesegeräte übertreffen. Das Problem der veraltenden Trägermedien ist im Prinzip geklärt: man kopiert die Daten automatisch (mittels Robotern) auf die jeweils aktuelle Trägergeneration. Die logistischen Herausforderungen sind dennoch nicht unbeträchtlich. Wohl organisierte Repositorien finden ihre Daten dann auch am neuen Speicherort wieder; für die Nutzer sind die Vorgänge vollkommen transparent. Kopiergeschützte CD-ROMs hingegen fallen bei solchen Prozessen unter den Tisch.

Die viel größere Herausforderung stellt die Interpretierbarkeit der Dokumente bzw. ihre Darstellbarkeit dar. So kann selbst Word2010 nicht mehr eine Publikation einlesen, die mit einem frühen Vorläufer von Word2010 erstellt wurde. Man könnte glauben, dass solche Schwierigkeiten durch permanente aktive Dokumentenmigration in die jeweils aktuelle Darstellungsform gelöst werden könnten - verbunden mit der Hoffnung, dass das Ergebnis 100% mit dem Original übereinstimmt. Dieser Ansatz versagt jedoch völlig, wenn die gesamte Laufzeitumgebung notwendig ist, um ein Ergebnis darstellen zu können. Hier hilft nur die Emulation, also die softwaremäßige Aufbereitung der Vergangenheit auf aktuellen Rechnern. Gerade zu diesem Thema laufen in Freiburg, im Rahmen des EU-Projekts PLANETS, aktuelle Forschungsarbeiten mit sehr erfolgversprechenden Ergebnissen.

Es zeigt sich, dass auf die Betreiber von Repositorien wesentlich höhere Anforderungen als zu Beginn erwartet zukommen, wenn sie ihre Daten auch für die Zukunft nutzbar halten wollen. Diese Anforderungen gehen mit Betriebskosten und Betriebskompetenz einher. Werden die oben aufgezeigten Transportketten in der Zukunft nur einmal unterbrochen, dann sind die gespeicherten Daten nur noch Datenmüll, weil niemand sie mehr zu interpretieren vermag.

Wenn Repositorien aber diese Herausforderung meistern, so werden sie unverzichtbare Grundpfeiler der Informationsvermittlung, die das Wissen in die fernsten Ecken der vernetzten Welt liefern können - ähnlich wie der Buchdruck einst zur Wissensverbreitung beitrug.