

Wissen nutzen statt Papier sortieren: Entwicklungstrends in Pressearchiven

Referat am SFJ-Herbstseminar "Archivierung - Wege aus dem Chaos"

Josef Wandeler, Trialog AG

Sehr geehrte Damen und Herren

Dieses Seminars widmet sich hauptsächlich der Frage, wie das persönliche Archiv mit vernünftigem Aufwand organisiert werden kann und soll. In meinem Referat will ich einen anderen Aspekt des Thema "Archive" beleuchten: die Pressearchive und -dokumentationen der Verlage, die auch ein wichtiges Instrument der journalistischen Arbeit sind. Ich möchte Ihnen kurz skizzieren, welche Trends die Entwicklung in diesem Bereich prägen und welche Chancen und Möglichkeiten sich für die Verlage hier eröffnen.

Die Vergangenheit: das Papierarchiv

Die Bezeichnungen *Pressearchive* und *Pressedokumentation* werden heute praktisch synonym verwendet; früher hat man eher von Archiven gesprochen, seit rund 20 Jahren hat sich der Begriff Dokumentation durchgesetzt. Es macht aber durchaus Sinn, die beiden Begriffe zu unterscheiden:

Genauso wie jedes Unternehmen ein Archiv hat, wo Geschäftsdokumente abgelegt werden, hat auch jeder Verlag ein Archiv, wo u.a. die produzierten Publikationen gesammelt werden. Dies konnte in der traditionellen Arbeitsweise auf zwei verschiedene Arten getan werden:

Die einfachste Art eines Pressearchiv ist natürlich, die einzelnen Publikationen zu sammeln und chronologisch geordnet aufzubewahren. Genau dies tut auch jede Bibliothek, die Zeitschriften oder Zeitungen sammelt und sie den Benutzern zur Verfügung stellt. Allerdings ist diese Art der Archivierung nur sehr begrenzt nutzbar: Eine thematisch orientierte Suche nach einzelnen Artikeln ist nur möglich, wo entsprechende Register produziert wurden und auch diese umfassen selten mehr als einen Jahrgang - solange das Erscheinungsdatum eines Artikels nicht bekannt ist, muss mit entsprechendem Zeitaufwand Jahrgang für Jahrgang einer Zeitschrift durchgeblättert werden. Das mag akzeptabel sein für Wissenschaftler mit viel Zeit - nicht aber für Journalisten, die unter Zeitdruck arbeiten und einen Redaktionsschluss einhalten müssen.

Wenn eine gezielte thematische Suche nach einzelnen Artikel möglich sein soll, müssen die Artikel auch einzeln nach Themen geordnet abgelegt werden. Dies ist natürlich wesentlich arbeitsaufwendiger, da jeder Artikel ausgeschnitten werden muss, allenfalls auf ein einheitlichen A4-Format umbrochen wird und dann in eines oder mehrere Themen-Dossiers abgelegt wird. Der Hauptaufwand dabei ist nicht die manuelle Arbeit, sondern der intellektuelle Aufwand, denn für die Dossiers muss eine nach Sachthemen gegliederte Systematik erstellt und laufend gepflegt werden und die Artikel müssen von fachlich kompetenten Leuten den richtigen Dossiers zugeordnet werden. Solche Pressearchive sind vor allem in den grossen Verlagshäusern entstanden: ein typischer Vertreter dafür ist das Redaktionsarchiv der NZZ, wo seit Jahrzehnten die eigene Zeitung ausgewertet und die Artikel in Dossiers abgelegt wurden.

Eine Erweiterung dieser Art Pressearchive sind die Pressedokumentationen. Sie unterscheiden sich dadurch, dass neben den hauseigenen Publikationen auch fremde Quellen ausgewertet werden. Damit wird ein wesentlich breiteres Informationsangebot aufgebaut und eine

zeitungsübergreifende Recherche ist möglich. Solche Pressedokumentationen sind v.a. in den grossen Medienunternehmen entstanden, wo die nötigen Mittel für diese aufwendige Arbeit zur Verfügung standen. In der Schweiz waren es der Ringier-Verlag und Fernsehen DRS, die seit den 70er-Jahren die grössten Pressedokumentationen aufgebaut haben. Der in unserem Sprachraum berühmteste Vertreter dieser Gattung ist sicher das geradezu legendäre Archiv des Spiegel in Hamburg. Infolge ihrer Grösse bezüglich der Menge der gespeicherten Artikel wie auch der Breite des inhaltlichen Spektrums, die sie abdecken, wurden hier auch sehr umfangreiche und teilweise hochkomplexe Systematiken aufgebaut. Beim Spiegel umfasste die frühere Papierdokumentation über 50'000 Dossiers zu einzelnen Themen, in anderen Verlagen war die Situation ähnlich. Es versteht sich von selbst, dass nur Spezialisten überhaupt in der Lage waren, mit diesen Dokumentationen zu arbeiten, sowohl bei der richtigen Zuordnung der Artikel wie auch bei der Recherche.

Die Gegenwart: das elektronische Archiv

Soviel zur Vorgeschichte. In den vergangenen rund zwanzig Jahren hat sich die Situation drastisch verändert und das Stichwort dazu heisst *Digitalisierung*. Die Umstellung auf elektronische Produktionsmethoden hat im Bereich Satz und Druck angefangen, sich dann auf die Redaktionen ausgedehnt und schliesslich mit etwas Verzögerung auch die Archive und Dokumentationen erreicht.

Für die Pressearchive bedeutet dies, dass die elektronisch produzierten Artikel direkt aus dem Redaktionssystem in eine Volltextdatenbank transferiert und den Nutzern für Recherchen angeboten werden können. Diese Umstellung hat zwei entscheidende Verbesserungen gebracht: Erstens ermöglicht eine Datenbank eine sehr viel komfortablere Abfrage nach den unterschiedlichsten Suchkriterien; die Recherche ist nicht mehr an die monohierarchische Struktur der Ablagesystematik gebunden. Und zweitens ist eine Datenbank nicht standortgebunden; sie kann intern direkt auf dem reaktionellen Arbeitsplatz und extern einem weiteren Benutzerkreis angeboten werden. Seit vielen Jahren sind solche elektronischen Pressearchive über Datenbankanbieter wie Dialog gegen Bezahlung zugänglich und die Verlage haben hier eine zusätzliche Einnahmequelle erschlossen. Mit dem Siegeszug des Internet ist diese Zweitverwertung noch viel wichtiger geworden und aus dem einstigen "Abfallprodukt" sind eigenständige Publikationsformen auf CD-ROM oder als Online-Ausgaben entstanden.

Der gleiche Prozess hat auch in den Pressedokumentationen stattgefunden, wo die aus Fremdquellen ausgewählten Artikel elektronisch gespeichert werden. Eine Pionierrolle hat dabei die Pressedokumentation des Verlags Gruner+Jahr in Hamburg gespielt: Bereits in den 70er-Jahren wurde dort begonnen, Zeitungsausschnitte zu scannen, als elektronische Kopie auf Bildplatten zu speichern und über eine Datenbank mit formalen und inhaltlichen Daten zu erfassen und zugänglich zu machen. Eine Pionierrolle spielte Gruner+Jahr auch in einem zweiten Punkt, indem die Pressedatenbank für externe Kunden zugänglich gemacht und zu einem eigenständigen kommerziellen Geschäft wurde. In den letzten Jahren haben viele grosse Pressedokumentationen im deutschen Sprachraum diese Umstellung von der Papierablage zur Datenbank vollzogen oder sind daran, entsprechende Projekte zu realisieren. Dabei haben sich ganz unterschiedliche Vorgehensweisen, wie die Artikel aus fremden Quellen digitalisiert werden, herauskristallisiert:

a) Die gedruckte Vorlage wird gescannt, der einzelne Artikel wird via cut&paste ausgeschnitten und ins Format A4 umbrochen und als Bilddatei gespeichert. In der Datenbank wird diese Bilddatei mit formalen Angaben (Name der Quelle, Ausgabe, Seite, Autoren etc.) und Schlagwörtern, die den Inhalt kennzeichnen, beschrieben. Diese Angaben sind die Suchkriterien, nach denen recherchiert werden kann und als Ergebnis der Recherche werden die Faksimiles der Artikel angezeigt.

b) Eine weitere Variante besteht darin, die Bilddatei mit dem Artikel mit einer Texterkennungs-Software wieder in maschinenlesbaren Text zu verwandeln und in der Datenbank neben den formalen und inhaltlichen Angaben auch den Volltext des Artikels zu speichern. Damit wird erreicht, dass einerseits eine Freitextrecherche über den gesamten Text möglich ist und als Ergebnis der Recherche entweder der Text in maschinenlesebarer Form oder das Faksimile angezeigt werden kann.

Beide Varianten haben aber den Nachteil, dass der maschinenlesbare Text mit entsprechendem Aufwand aus einem Printprodukt erzeugt wird, das ursprünglich bereits elektronisch vorhanden war. Inzwischen haben sich Firmen darauf spezialisiert, genau diese Arbeit als externe Dienstleister anzubieten.

c) Die dritte Variante verzichtet auf diesen Umweg: Gegen Bezahlung werden die Daten der kompletten Publikationen aus dem Redaktionssystem des anderen Verlages bezogen, konvertiert und in die eigene Datenbank importiert. Dieses Verfahren ist - gerade bei grossen Mengen - wesentlich schneller und weniger arbeitsaufwendig. Dafür wird auf das Faksimile der Artikel verzichtet, was anfänglich als grosser Nachteil bewertet wurde. Inzwischen können die Artikel dank Internet- bzw. Intranet-Technologie als Quasi-Faksimile angezeigt werden, das weitgehend dem Layout der Online-Ausgaben der Originalpublikationen entspricht.

Mit dem Übergang vom Papierdossier zur Volltextdatenbank verändern sich auch die Möglichkeiten für inhaltliche Recherchen. Solange in einer elektronischen Pressedatenbank die Artikel nur als Bilddateien gespeichert sind, müssen die Inhalte der einzelnen Artikel mit einer ausführlichen und aufwendigen Beschlagwortung erfasst werden. Bei Gruner+Jahr wurde dafür eine sehr umfangreiches und hochkomplexes Vokabular verwendet, das nur von spezialisierten Dokumentalisten benutzt werden konnte. Seit bei Gruner+Jahr neben den Bilddateien mit den Faksimiles auch die Volltexte der Artikel gespeichert sind und durchsucht werden können, zeigte sich bald, dass eine so komplexe Beschlagwortung letztlich nur ein Hindernis darstellt. In den vergangenen Jahren ist deshalb die aufwendige Systematik ersetzt worden durch einen wesentlich schlanken Schlagwort-Thesaurus. Damit ist die Recherche durch Endbenutzer aus den Redaktionen wesentlich erleichtert worden und diese können einen grossen Teil ihrer Datenbankabfragen auch ohne Hilfe eines Recherchedokumentalisten durchführen. Auf der Input-Seite ist der Aufwand für die inhaltliche Erschliessung aber immer noch erheblich, da jeder einzelne Artikel nach verschiedenen Kriterien beschlagwortet werden muss.

Mit verschiedenen Ansätzen wird versucht, diesen Aufwand zu vermindern oder möglichst ganz zu vermeiden. Der radikalste Weg ist natürlich der völlige Verzicht auf eine intellektuelle Beschlagwortung, im Vertrauen darauf, dass mit den Mitteln der Freitextrecherche ja jedes Wort suchbar ist und mit den weiteren Möglichkeiten wie logische Verknüpfung von Suchbegriffen, Kontextoperatoren etc. so jedes Thema eingegrenzt werden könne. In der Praxis zeigen sich rasch die Grenzen dieses Verfahrens, da die sprachliche Vielfalt dazu führt, dass derselbe Sachverhalt sich in ganz unterschiedlichen Worten und Wortformen ausdrückt. Eine präzise Suche, die auch alle relevanten Dokumente findet, wird da schwierig und schnell einmal aufwendig. Dieses Problem lässt sich allerdings vermindern, wenn die gespeicherten Volltexte linguistisch aufbereitet werden, indem z.B. alle einzelnen Wörter auf ihre grammatikalische Grundform zurückgeführt und die im Deutschen sehr häufiger Mehrwortbegriffe in ihre einzelnen Bestandteile zerlegt werden. Mit diesem System wird in der Datenbank der Deutschen Presseagentur in Hamburg gearbeitet, zudem werden die einzelnen Agenturmeldungen mit einer rudimentären Klassifikation von ca. 40 Begriffen einer Themengruppe zugeordnet. Auf diese Weise erreichte man, dass mit einem minimalen Aufwand von intellektueller Erschliessung zufriedenstellende Rechercheergebnisse erzielt werden können.

Ein ganz anderer Ansatz sind die Verfahren zur automatischen oder maschinengestützten Indexierung, die gegenwärtig entwickelt werden und an verschiedenen Orten bereits im produktiven Einsatz sind. Auch diese Systeme basieren auf einem Erschliessungs-Vokabular, einer Systematik oder einem Schlagwort-Thesaurus, das intellektuell entwickelt und auch

laufend gepflegt werden muss. Hier sind es aber nicht Dokumentalisten, die jedem Artikel die entsprechenden Begriffe zuweisen, sondern ein Programm, das den Text eines Artikels analysiert und daraus das Thema erkennt und die richtigen Begriffe zuweist. Der intellektuelle Aufwand wird damit nicht ganz abgeschafft, aber er wird kleiner und verlagert sich: die Aufgabe der Dokumentalisten konzentriert sich darauf, eine laufende Qualitätskontrolle der Ergebnisse vorzunehmen und das System permanent zu trainieren, damit es neue Themen und Sachverhalte richtig erkennen kann.

Die Zukunft: Kooperation

Damit sind wir schon bei der Zukunft der Pressearchive und Pressedokumentationen angelangt. Oft fragt man sich, ob die heutigen grossen Pressedokumentationen nicht ihre Existenzberechtigung verlieren werden, wenn bald alle Publikationen elektronisch auf dem Internet verfügbar sind. Es sind vor allem zwei Gründe, warum dies wohl nicht der Fall sein wird. Erstens besteht die Qualität von Pressedokumentationen gerade darin, dass sie aus der Fülle vorhandener Dokumente eine gezielte Auswahl treffen, die dem Informationsbedarf ihrer Nutzer angepasst ist und so eine schnellere und präzisere Suche ermöglichen. Und zweitens haben sie den Vorteil, dass mit einer Suchabfrage in einem Bestand eine Vielzahl von Quellen abgefragt werden kann. Die grossen Suchmaschinen auf dem Internet bieten diese Möglichkeit zwar auch, sie liefern aber - mindestens heute und in absehbarer Zukunft - auch immer eine erhebliche Menge an Ballast. Genau dies ist auch der Vorteil einer grossen Pressedatenbank gegenüber den CD-ROM-Ausgaben einzelner Publikationen: So nützlich diese auch sind - sobald Artikel zu einem Thema unabhängig von der Quelle gesucht werden, ist die Recherche in einer Vielzahl von CD-ROM-Archiven doch wesentlich aufwendiger als eine einzige Recherche in einer Datenbank.

Ein ganz zentrales Stichwort für die Zukunft der Pressedokumentationen heisst *Kooperation*. Denn in allen Pressedokumentationen, die ein breites Spektrum von Quellen berücksichtigen, sind es teilweise genau dieselben Publikationen, die ausgewertet werden. Im deutschen Sprachraum sind dies die grossen überregionalen Zeitungen sowie auch die wichtigsten internationalen Leitblätter. Solange die Pressedokumentationen in Papierform erstellt wurden, waren sie standortgebunden und damit war diese Mehrfacharbeit nicht zu umgehen. Der schnelle Zugriff auf eine elektronische Pressedokumentation ist aber nicht mehr distanzabhängig und was an einem Ort getan wird, kann auch entfernten Nutzern zugänglich gemacht werden.

Aus diesem Grund ist es naheliegend, dass Pressedokumentationen kooperieren und Mehrfacharbeiten eliminieren, indem sie ihre Auswertung in einen gemeinsamen Topf - sprich Pressedatenbank - einfliessen lassen, der dann allen Beteiligten zur Verfügung steht. In der Schweiz ist so die Schweizer Mediendatenbank entstanden als gemeinsame Pressedatenbank der Verlage Ringier, TA-Media sowie Fernsehen DRS. In Deutschland haben verschiedene ARD-Anstalten (WDR, NDR, SWR u.a.) damit begonnen, eine gemeinsame Pressedatenbank aufzubauen und damit verhindert, dass mehrere identische Projekte gleichzeitig realisiert wurden. Das Prinzip solcher Kooperationen besteht meist darin, dass die Auswertung der gemeinsam genutzten Quellen unter den Partnern aufgeteilt wird und daneben jeder Partner zusätzlich weiteres Material einbringen kann - in der ARD-Kooperation z.B. regionale Publikationen, die nur für das eigene Haus relevant sind.

Die Problematik solcher Kooperationen besteht meistens darin, dass unterschiedliche Traditionen und Arbeitsweisen zusammengeführt und vereinheitlicht werden müssen. Dies betrifft insbesondere die Selektionskriterien, nach denen einzelne Artikel für die Auswertung ausgewählt werden und eine einheitliche Inhaltserschliessung mit einem gemeinsamen Vokabular.

Eine andere Form der Kooperation besteht darin, dass verschiedene Verlage ihre Volltextarchive mit den hauseigenen Publikationen in einen gemeinsamen Pool einbringen und

diesen vermarkten. Ein Beispiel dafür ist die 1999 gegründete *Pressemonitor GmbH* *, in der die grossen deutschen Verlage (FAZ, Gruner+Jahr, Handelsblatt, Springer, Burda, Spiegel, etc.) beteiligt sind. Ziel ist es, in einer gemeinsamen Datenbank den Kunden gegen Bezahlung aktuelle Artikel für die Produktion hausinterner Pressespiegel anzubieten. Damit versuchen diese Verlage das Geschäft, das bisher spezialisierte Firmen wie Argus betrieben haben, selber in die Hand zu nehmen und sich damit eine neue Einnahmequelle zu erschliessen. Ob dieses Projekt auf dem Markt erfolgreich sein kann, wird vor allem davon abhängen, ob sich neben den grossen Verlagen auch die Verlage der kleinen und mittleren Zeitungen und der Fachzeitschriften anschliessen, damit den Kunden ein ausreichend grosser Fundus angeboten werden kann.

Damit ist auch die kommerzielle Seite der Pressedatenbanken angesprochen. Das grosse Geschäft lässt sich mit ihnen zweifellos nicht machen - im Gegensatz zu Datenbanken mit technischen, naturwissenschaftlichen oder Finanzinformationen. Die Erfahrungen zeigen, dass die Bereitschaft der potentiellen Kunden, für Presseinformationen zu bezahlen, mit steigendem Preis rapide zurückgeht. Die Vermarktung von Pressedatenbank über Datenbankanbieter oder via Internet ist denn primär ein Beitrag an die Gemeinkosten für die Auswertung. Als Beispiel seien hier die Preise der Schweizer Mediendatenbank für Einzelkunden, die über Internet recherchieren, angeführt:

- CHF 6.50 pro Volltext-Artikel (v.a. Publikationen von Ringier und TA-Media).
- CHF 2.50 pro Referenz-Dokument (bibliografische Informationen und Inhaltserschliessung zu einem Artikel).
- CHF 4.50 pro Faksimile-Kopie (neuere Referenz-Dokumente sind generell mit einer Faksimile-Seite verknüpft.).
- CHF 50.- als Mindestverrechnung pro Monat.

Als Vergleich dazu die Preisstruktur von Genios, einem der führenden Datenbankanbieter für deutschsprachige Pressedatenbanken:

- DM 9.- Mindestverrechnung pro Monat
- DM 2.- pro Suchanfrage inkl. Anzeige der Trefferliste
- Vollanzeige pro Artikel:
 - DM 3.60 - 4.80 für Zeitungsdatenbanken
 - DM 4.20 - 5.40 für Fachzeitschriften
 - DM 3.00 - 6.60 für Firmenadressen
 - DM 6.60 - 19.20 für Firmenprofile
 - DM 30.00 - 78.00 für Bilanzdatenbanken/Firmenreports

Copyright © 1999, Trialog AG

*) vgl: Georg Wallraf: "Pressemonitor GmbH - eine Lösung für den elektronischen Pressespiegel", in: INFO7, Heft 2 1999, Baden-Baden, Nomos-Verlag, 1999.