

Die Archivierung der Webinhalte im Westdeutschen Rundfunk

Chancen für multimediales Produktionsarchiv von WDR Online

Vom Hanno Jochemich



Hanno Jochemich
IID der Fachhochschule
Potsdam
Information Specialist
heute WDR Köln
Eigelstein 106-108
50668 Köln
0152 292 696 35
Hanno.Jochemich@
wdr.de

Schönen guten morgen zusammen! Vielen Dank für den Newcomer-Preis und die Möglichkeit, Ihnen hier mein Thema vorstellen zu können. Es handelt sich um das Ergebnis eines 30seitigen Aufsatzes, den ich als Abschluss meines Volontariats zum Wissenschaftlichen Dokumentar am Institut für Information und Dokumentation in Potsdam eingereicht habe.

Hier ein Überblick zu den nächsten 20 Minuten: Zunächst möchte ich Ihnen erklären, warum es hier überhaupt geht. In Punkt. 2 und 3 möchte ich erzählen, wer bei WDR Online wie welche Inhalte produziert und natürlich was wir als Archivare damit anfangen können. Bei 4. bespreche ich die Chancen für ein multimediales Produktionsarchiv anhand von 3 Optionen und gebe eine Empfehlung ab. Zum Schluss gibt es einen kurzen Ausblick und ein Fazit.

1 Zur Einstimmung, in welcher Größenordnung wir uns bewegen, habe ich Ihnen eine WDR Online Statistik mitgebracht. Im Februar 2013 wurden die Seiten von WDR Online 22,5 Millionen mal besucht – das heißt ca. 800.000 mal täglich – und dabei 105 Millionen mal angeklickt. Wir haben es hier also mit einem Verbreitungsweg des WDR zu tun, der ziemlich viele Menschen erreicht. Deshalb sage ich einerseits das, was eh längst klar ist: Das Internet hat sich neben TV und Hörfunk längst etabliert. Und das spiegelt sich auch in den Online-Inhalten wider: Der Internetauftritt des WDR hat sich – so wie bei allen Öffentlich-Rechtlichen Rundfunkanstalten – von der Programmbegeleitung zur eigenständigen Plattform für Information, Bildung, Beratung und Unterhaltung entwickelt. Webinhalte sind auch Programmvermögen.

Andererseits beschäftigen die Archive sich seit rund 10 Jahren mit der Webarchivierung, ohne dass es außer einzelnen Initiativen eine grundlegende Strategie zur Thematik gibt. Scheinbar kam immer etwas Wichtigeres dazwischen. Dabei ist der Auftrag klar: „Die Abteilung Dokumentation und Archive ist nun gebeten, ein Konzept zur Archivierung erhaltenswerter Online- Publikationen zu erstellen.“ (Regelung zur Depublikation von Online-Beiträgen im WDR). Die

Archivierung des Online-Programmvermögens ist ein Job für die Archive.

- Für die Archive stellen sich im Zusammenhang mit den Online-Inhalten einige Fragen:
- Wie kann das Programmvermögen nachhaltig archiviert werden?
- Für wen könnten die Webinhalte bereitgestellt werden? Und wie soll das passieren?
- Bilden Online-Quellen einen Mehrwert für die laufende Produktion? Welche Varianten zur Archivierung sind möglich?
- Welche Chancen bietet ein multimediales Produktionsarchiv für Webinhalte? Welche Rolle nimmt das Archiv im Prozess der Online-Produktion überhaupt ein?

Prinzipiell gibt es bei den Öffentlich-Rechtlichen zwei Konzepte zur Webarchivierung: Konzept Nr. 1 ist die Historische Webarchivierung. Dabei handelt es sich im übertragenen Sinne um einen Sendemitschnitt (das ist dann ein sogenanntes Snapshotarchiv). Die Archivierung dient in erster Linie zur Unternehmensdokumentation. Uns soll aber heute vielmehr das 2. Konzept interessieren, nämlich die Produktionsarchivierung von Webinhalten. Dabei geht es darum, die Webinhalte für Fernseh-, Radio- und Internet-Redaktionen wiederverwertbar zu machen.

2 Punkt 2 widmet sich der Produktion von WDR Online. Das ist ein ganz kurzer organisatorischer und technischer Überblick. Im gesamten WDR Onlineangebot fließen Informationen aus verschiedenen Bereichen zusammen. WDR Online ist mit über 70 Planstellen und zahlreichen freien MitarbeiterInnen ausgestattet. Der Programmbereich Internet nimmt dabei die zentrale Rolle ein und verantwortet vor allem wdr.de, sportchau.de und die WDR-Mediathek. Daneben können die Redaktionen aus Hörfunk und Fernsehen bei WDR Online ihre Themen und Sendungen breiter darstellen und diskutieren.

Zur Online-Produktion sind im WDR mehrere CMS im Einsatz. Das macht die Analyse nicht gerade einfacher, aber seit 2011 findet auch eine zunehmende

*Vortragsmanuskript
(gehalten auf der
Frühjahrstagung des
vfm am 23. April
2013)

Vereinheitlichung dieser heterogenen Produktionsumgebung statt. Die Redaktionen steigen nach und nach auf das neue CMS Sophora um.

Bei der Produktion von Webinhalten im CMS spielen 3 Komponenten eine Rolle: Erstens geht es um den Content, also die eigentlichen Inhalte, die Essenzen. Zweitens geht es beim Design darum, wie dieser Content bearbeitet und dargestellt wird, also Schrifttyp, Logos etc. Und drittens wird mit dem Layout bestimmt, wie alle Einzelelemente auf der Website anzuordnen sind.

Für uns ist hier wichtig, dass diese 3 Komponenten unabhängig voneinander bearbeitet und weiterentwickelt werden können. Denn zur Wiederverwendung in der Produktion ist der Content entscheidend. Denn der Content ist die Summe aus Essenz und Metadaten – und das sind die Elemente, die auch für die Produktion in Fernsehen und Radio relevant sind.

Bei der Online-Produktion gibt es aus Sicht der Archive einige spezielle Probleme: Das Verweildauer-Konzept sorgt dafür, dass die meisten Webinhalte nach einer gewissen Zeit depubliziert werden müssen. Bei so einem „toten Link“ ist wahrscheinlich jeder schonmal gelandet. Hier entsteht vor allem für diejenigen ein Problem, die zwar Interesse an den Inhalten, aber keinen Zugriff auf das CMS haben.

Ein anderes Problem ist das praktizierte Speicherkonzept. Theoretisch speichert das CMS zwar alles für immer, aber praktisch sieht die Lage etwas unübersichtlicher aus: Die Pflicht zur 3-monatigen Beweissicherung wird eingehalten, aber darüber hinaus kostet Speicherplatz vor allem Geld. Die Speicherzeit auf dem CMS, Produktionsservern oder in anderen Mediendatenbanken beträgt oft – wenn überhaupt – nur wenige Jahre.

Ein weiteres Problem ist der Relaunch eines neuen CMS. Der Wechsel auf ein neues CMS geschieht oft unter Termindruck und führt bei der Datenmigration zu Schnellschüssen. Die Online-Produktion ist generell ziemlich schnelllebig und diese Systemwechsel von einem CMS zu einem anderen entwickeln sich eher zu einem latenten Dauerthema anstatt einmalig aufzutreten.

Zusammengefasst heißt das, dass die ehemaligen Fensterbank-Archive der Fernsehleute sich zu Festplatten-Archiven im Onlinebereich geworden sind. In den Zeitdimensionen der Archive sind Online-Inhalte eine kurzlebige Quelle. Datenverluste sind momentan nicht zu verhindern.

3 Im nächsten Punkt geht es um die Inhalte – und die damit verbundenen Fragen:

Was machen die da eigentlich? (Das ist die inhaltliche Analyse.) und Brauchen wir das? (Das ist die archivarisches Analyse.) Ein kurzer Blick auf die Besonderheit von Webinhalten: Webinhalte zeichnen sich in erster Linie durch ihre Multimedialität aus. Sie sind ein nicht-linearer Mix aus den klassischen drei Medien Fernsehen, Hörfunk und Printmedien. Und nicht nur das: Denn darüber hinaus werden die

klassischen Inhalte zu neuen Medientypen wie Webmagazin, Webspecial, Webdokumentation etc. verschmolzen. Die Kreativität der Onliner bietet hier interessanterweise immer neue Formate.

Mit Blick auf die Dokumentation der Webinhalte sollte man als Medienarchivar zwischen Massencontent und Premiumcontent unterscheiden. Das heißt, dass ich mich bei der Inhaltsanalyse fragen muss, welche Teile von WDR Online lediglich ein neu verpackter Bausatz aus TV- und Radio-Inhalten sind oder in welchem Maß hier auch eigenständige journalistische Leistung geboten wird – sprich Premium-Content bzw. auch genuiner, originärer oder Online-Only-Content genannt. Denn für die Produktionsarchivierung macht es keinen Sinn, sich mit den Medienobjekten zu beschäftigen, die bereits in den Archivdatenbanken vorhanden sind.

Konzentrieren wir uns also auf die originären Webinhalte: Was gibt es da bei WDR Online? Der Löwenanteil der originären Webinhalte besteht aus Texten. Dabei handelt es sich um Interviews, Porträts, Hintergrundberichte oder Analysen – oft mit klarem NRW-Bezug –, deren Inhalt sich nicht unbedingt in der Presse wiederfinden lässt.

Die Blogs liegen zum Teil auch vertont als Audiofiles vor. Viele Fotos von WDR Online werden oft zum ersten Mal oder ausschließlich auf den Online-Seiten publiziert. Außerdem gibt es zahlreiche grafisch bearbeitete Fotos, wie etwa Bildergalerien. Eine andere Quelle für Fotos sind Aktionen wie das Foto- und Videoprojekt „digit“, bei dem der WDR Material von seinem Publikum erhält. Die Menge an genuinem Videomaterial fällt etwas geringer aus.

Zu Events erstellen Radiowellen Videoblogs oder kurze Beiträge. In seinem 1Live-Fernseher veröffentlicht z.B. 1Live einige originäre Beiträge. Die Sportschau veröffentlicht einige Kurzreportagen und Beiträge im Netz. Es gibt Langversionen von Interviews und Berichten, die nur online publiziert werden. Originäre Audios findet man bei WDR Online übrigens fast gar nicht. Sie tauchen höchstens im Zusammenhang mit Webspecials auf. Die webspezifischen Inhalte sind selbstverständlich originäre Quellen. Hier gibt es zum Beispiel Webmagazine, Panorama-Animationen oder Infoboxen. Grundsätzlich sind sich eigentlich alle, die sich mit dem Thema beschäftigen, einig, dass die Produktion originärer Webinhalte weiter zunehmen wird.

4 Als Archivar frage ich mich jetzt: Wie bekommen wir diese originären, einmaligen Inhalte ins Archiv? Wie stehen also die Chancen für ein multimediales Produktionsarchiv? Und daran anknüpfend: Wie können wir die Inhalte wieder für das Programm bereitstellen? Prinzipiell soll ein Produktionsarchiv einen Mehrwert für die aktuelle Programmproduktion bieten und zur effizienten Wiederverwertung des Programmvermögens beitragen. Ich habe 3 Optionen für ein Produktionsarchiv von WDR Online durchgedacht, die wir uns jetzt mal anschauen.

Option A ist das multimediale Produktionsarchiv – die naheliegendste Option. Dabei geht es um ein paralleles Archivsystem innerhalb des CMS. Das CMS Sophora, das bei WDR Online die zentrale Rolle spielt, lässt sich beliebig erweitern – auch um eine sogenannte Archiv-Komponente. Hier könnten depublierte, archivierungswürdige Inhalte geparkt, dokumentiert und wiedergefunden werden. Die Archivare engagieren sich hier direkt bei WDR Online.

Die Vorteile liegen auf der Hand: Die technische Infrastruktur für so eine Lösung ist bereits vorhanden. Das Archiv ist nach an der Online-Produktion und bleibt am Ball. Sämtliche Medientypen können problemlos komplett und erschlossen und vorgehalten werden, solange das CMS das kann.

Der größte Nachteil ist bei dieser Option die Abhängigkeit vom CMS. Bisher lässt sich nicht vorhersagen, wie performant das System mit immer mehr Inhalten und Usern bleiben wird. Es gibt da bisher keinen vergleichbaren Fall zum WDR. Die Frage der Datenmigration ist bei der Arbeit im CMS nur aufgeschoben. Es handelt sich um eine verhältnismäßig kurzfristige Schnelllösung – die Langzeitarchivierung ist nicht geplant.

Als Nutzer kommen primär die Onliner infrage, und die haben bisher kein besonderes Interesse an der Dokumentation und Archivierung ihrer Inhalte geäußert. Noch haben sie einen Durchblick auf dem CMS – bleibt abzuwarten, wie das zukünftig aussieht. Für alle anderen Programmschaffenden wäre das CMS eine weitere lästige Rechercheplattform. Lohnt es sich dafür also, einen neuen Archiv-Workflow zu implementieren? Diese Lösung kann wahrscheinlich nicht mit anderen Systemen außerhalb des WDR verbunden werden. Option B ist ein dezentrales Produktionsarchiv. Was heißt das? Bei dieser Variante werden die originären Webinhalte auf die vorhandenen Archivdatenbanken verteilt. Dazu benötigt man vor allem Schnittstellen vom CMS, Produktionsservern und Mediendatenbanken ins Archiv: Es würde sich um einen automatischen und in Ausnahmefällen (z.B. bei externen Produktionen) um einen manuellen Import des Online-Contents handeln. Größter Vorteil bei dieser Option ist die bestehende Archiv-Infrastruktur. Das Archiv agiert auf eigenem Terrain und behält die Hoheit über Inhalte und Workflows. Und so weit die Archivdatenbanken ARD-weit vernetzt sind, lassen sich die Inhalte dann auch in den anderen Rundfunkanstalten finden.

Der Nutzerkreis entspricht dem aktuellen Nutzerkreis der Archive. Wir brauchen hier eigentlich keine zusätzlichen Rechercheoberflächen, es sei denn – und damit komme ich zu den Nachteilen – wir finden keine passable Lösung, die webspezifischen Inhalte zu archivieren. Ein großer Nachteil beim dezentralen Produktionsarchiv besteht darin, dass hier die monomediale Archivlandschaft auf eine multimediale Produktionsumgebung trifft, mit der sie nicht umgehen kann.

Wir müssen also klären, wo wir multimediale Objekte am ehesten unterbringen würden, wenn es

nicht noch eine Datenbank dafür geben soll. Und wie wollen wir den Verlust der Zusammenhänge verhindern? Option C ist die Radikal-Lösung und heißt: Kein Produktionsarchiv. Das Archiv beschließt hier, sich soweit wie möglich aus dem Online-Bereich herauszuhalten. Hier wäre es denkbar, dass man interessierten Nutzerinnen und Nutzern, die nicht aus dem Online-Bereich kommen, entweder einen direkten Zugang zum CMS gibt. Denkbar wäre aber auch, dass man alle Webinhalte des CMS gesammelt durch eine Metasuchmaschine wie den Medienbroker oder über ein historisches Webarchiv zugänglich macht.

Die Vorteile liegen nahe: Das Archiv hat hier weniger zu tun. Die Inhalte sind trotzdem einigermaßen recherchierbar und zwar von allen, die Interesse daran haben. Der Nachteil hierbei ist, dass die Inhalte ungefiltert, doppelt und dreifach versioniert gefunden werden können. Und es ist die Frage, welche Nutzerin darin einen Mehrwert zu ihrer Recherche sieht. Es gibt wahrscheinlich ein Akzeptanzproblem gegenüber einer so oberflächlichen Recherche.

Meine Empfehlung für den WDR ist Option B, und zwar, weil sie pragmatisch ist und sich an den Bedürfnissen aller User orientiert. Sie ist überschaubar und wohl verhältnismäßig günstig. Die Inhalte sind so sicher wie alle anderen Inhalte der Archivdatenbanken auch. Und – wichtigster Punkt – diese Option ist ausbaufähig und kann verbessert werden. Das Problem Multimedialität vs. Monomedialität sehe ich als prototypisch für den Archivbereich der Öffentlich-Rechtlichen Medienhäuser an. Eine Verlinkung der Inhalte ist generell interessant und auch notwendig, denke ich.

5 Zum Abschluss noch ein Ausblick. Wer soll das eigentlich alles machen? Ich weiß, dass die Personalsituation nicht zum Besten steht, bin aber trotzdem der Meinung, dass sich ein rechtzeitiges Engagement im Online-Bereich lohnt – zumal die Lage inhaltlich, was den originären Content betrifft, noch halbwegs überschaubar ist. Das Archiv kann hier mit seinen Kernkompetenzen punkten. Ein Webarchivar oder ein Team Webarchiv könnte die wichtigsten Aspekte, die im Laufe des Vortrags deutlich wurden, bündeln.

Das inhaltliche Fazit lautet: WDR Online produziert originäre Webinhalte. Sie bieten einen produktionsbegleitenden Mehrwert für den Online-, Fernseh- und Hörfunkbereich. Bezüglich Produktionsarchiv für Webinhalte lässt sich Folgendes festhalten: Das Produktionsarchiv muss die originären Premium-Inhalte bereithalten. Das Produktionsarchiv muss für alle Programmschaffenden möglichst einfach erreichbar sein. Für den WDR ist die dezentrale Produktionsarchivierung momentan die beste Lösung.

Für die Archive lautet mein Fazit: Die Archive messen der Webarchivierung immer noch eine geringe Bedeutung bei. Sie verpassen dadurch den Anschluss an den „neuen“ etablierten Verbreitungsweg Internet. Ein Zügiges Engagement wird sich auszahlen!