

Barbara Lenk

## Webarchivierung musikwissenschaftlicher Internetressourcen /1/

Die Geschichte des Internets, wie es heute bekannt ist, begann am 6. August 1991: Tim Berners-Lee hat an der Europäischen Organisation für Kernforschung (CERN) die erste Website weltweit veröffentlicht. /2/ Der Begriff „Website“ lässt sich wie folgt definieren: „Eine Website ist ein virtueller Platz im World Wide Web, an dem sich meist meh-

rere Webpages oder Dokumente (Dateien) und andere Ressourcen befinden. Diese sind unter einer http-Adresse zu erreichen.“ /3/ In der Webarchivierung ist es wichtig, dass die Begriffe „Website“ und „Webpage“ korrekt verwendet werden. Unter einer „Website“ kann das gesamte Konstrukt einer Webpräsenz verstanden werden. Eine „Webpage“ ist eine einzelne Unterseite einer Website. Somit kann sich eine Website aus vielen verschiedenen Webpages zusammensetzen.

Websites haben sich vom Beginn des Internets bis in die Gegenwart stark verändert. Während

es zu Beginn textlastige Medien waren, sind sie heute multimedial und vereinen Text, Bild, Video, Audio und andere Inhalte. Im Rahmen einer Studie (Alexa Internet 2001) wurde die durchschnittliche Lebensdauer einer Website im Internet untersucht. Das Ergebnis ist ein Beleg für alle Aktivitäten der Webarchivierung: Websites haben eine Lebensspanne von 75 bis 100 Tagen. Damit drohen Informationen für die Forschung unwiderruflich verloren zu gehen. Internetressourcen können gelöscht, verändert oder an einer anderen Stelle platziert werden. Deshalb haben Bibliotheken, Archive und andere Einrichtungen damit begonnen, sich theoretisch und praktisch mit der Webarchivierung zu befassen.**/4/** Unter Webarchivierung versteht man alle Prozesse und Maßnahmen zur Auswahl, Erschließung, dauerhaften Speicherung, Archivierung sowie Bereitstellung von Websites für den Nutzer.**/5/** Das Ziel der Webarchivierung ist die dauerhafte Bewahrung von Websites und deren Inhalten, um diese u. a. als wissenschaftliche Quelle für die Forschung zu erhalten.

### Die Methode der Webarchivierung

Der Prozess der Datensammlung bei der Webarchivierung wird als „Harvesting“ bezeichnet. Das Ergebnis des Harvestings ist ein „Webschnitt“ (oder auch „Zeitschnitt“) einer Website. Dieser ist vergleichbar mit einem Schnappschuss der entsprechenden Website zum Zeitpunkt des Harvestings. Um eine Internetressource umfassend zu archivieren, geschieht ein solcher Prozess mehrmals zu festgelegten Zeitpunkten, z. B. täglich oder alle sechs Monate. Die aktive Durchführung der Datensammlung geschieht durch einen „Crawler“. Diese „Roboter“ (Software) folgen den weiterführenden Links innerhalb einer Website und analysieren wiederum die neu erkannten Links.**/6/** Dieses Verfahren wird beispielsweise bei Google und anderen Suchmaschinen eingesetzt. Die gecrawlten Websites werden in einem Archiv (z. B. Repositorium) abgespeichert. Vorher findet gege-

benenfalls eine Qualitätskontrolle statt. Als Speicherformate haben sich ARC bzw. WARC (XML-basiert) durchgesetzt. Der Zugriff auf ein Webarchiv kann über die WayBack Machine**/7/** oder andere Tools erfolgen.**/8/**

In der Literatur werden grundsätzlich drei Arten von Harvesting unterschieden: Domain Harvesting, Selective Harvesting und Event Harvesting. Im Rahmen des Domain Harvesting wird das gesamte Angebot an Websites einer Top-Level Domain archiviert, z. B. **\*\*\*.de**. Im Gegensatz dazu werden bei einem Selective Harvesting nur ausgewählte Websites gespeichert. Diese können etwa nach fachlichen Kriterien ausgewählt werden. Mit dem Event Harvesting werden wiederum alle Websites archiviert, die thematisch zu einem speziellen Ereignis gehören, beispielsweise Wahlen oder sportliche Großereignisse.**/9/**

### Grenzen der Webarchivierung

Aufgrund technischer und organisatorischer Gegebenheiten ist es nicht möglich, alle Websites zu archivieren. Mit Hilfe von Robot Exclusion Protokollen regeln die Betreiber einer Website eindeutig, wie die Crawler ihre Site bearbeiten dürfen. Die Websites können für Crawler gesperrt werden, es werden nur bestimmte Bereiche freigegeben bzw. es sind nur spezielle Typen von Crawlern erlaubt.**/10/**

Technische Besonderheiten von Websites erschweren ebenfalls die Aktivitäten der Webarchivierung. Ein großer Bereich ist in diesem Fall das Deep Web. „Dabei handelt es sich um Webseiten, die nicht statisch vorliegen, sondern basierend auf Anfragen dynamisch generiert werden (z. B. Telefonbücher, Kataloge, geographische Informationssysteme etc.).“**/11/** Diese Tatsache ist schwerwiegend, da das Deep Web einen deutlich größeren Umfang hat als das offen zugängliche Internet.**/12/** Webinhalte, die erst durch Authentifizierung oder ad-hoc-Abfragen generiert werden, können ebenfalls nicht archiviert werden. Eine dritte technische Grenze stellen dynamische Inhalte von Websites dar: Applikationen in

JavaScript, Flash-Anwendungen, YouTube-Videos u. a./13/ In allen diesen Feldern müssen (technische) Methoden und Möglichkeiten entwickelt werden, um die entsprechenden Webinhalte in Zukunft archivieren zu können.

#### *Rechtliche Aspekte*

Websites und deren Inhalte sind urheberrechtlich geschützt. Wenn ein Webarchiv frei und öffentlich sein soll, muss vor der Archivierung die Zustimmung des Seitenbetreibers eingeholt werden. Hierbei sollte ein besonderes Augenmerk auf Inhalten dritter Parteien liegen (Zustimmung muss vorhanden sein). Häufig sind Webarchive von daher als Dark Archive konzipiert. In diesem Fall ist der Zugriff nur in den Räumen der archivierenden Einrichtung, nicht von privaten Rechnern aus erlaubt./14/ Ein weiterer rechtlicher Aspekt ist der Schutz der Privatsphäre. Viele Inhalte im Internet können persönlicher Natur sein oder enthalten persönliche Daten und bedürfen daher eines besonderen Schutzes./15/

### **Webarchivierung in Musikbibliotheken am Beispiel der Virtuellen Fachbibliothek Musikwissenschaft**

Es gibt zahlreiche Websites mit musikwissenschaftlichen Inhalten: Websites von Komponistinnen und Komponisten,/16/ von musikalischen Spielstätten, musikalischen Gesellschaften, Bildungseinrichtungen aller Art, Blogs oder Wikis mit musikalischem Hintergrund und viele mehr. Charakteristisch ist die Heterogenität dieser Websites. Zum einen sind sie unterschiedlich gestaltet: textlastig bis multimedial. Zum anderen sind sie fachlich auf sehr unterschiedlichem Niveau: vom musikwissenschaftlichen Hintergrund bis hin zu einer rein laienhaften Darstellung.

Die Virtuelle Fachbibliothek Musikwissenschaft (ViFaMusik)/17/ ist ein zentrales Portal für Musikwissenschaft im deutschsprachigen Raum. Betreiber sind die Bayerische Staatsbibliothek (BSB)

und deren Projektpartner Staatliches Institut für Musikwissenschaft und Gesellschaft für Musikforschung. Das Portal ist Bestandteil des Fachinformationsdienstes Musikwissenschaft/18/ der BSB. Im Rahmen der ViFaMusik gibt es eine Kollektion musikwissenschaftlicher Websites.

Die fachliche Auswahl der Websites erfolgt durch die Verantwortlichen bei der ViFaMusik. Der technische Vorgang der Webarchivierung wird durch das Bibliothekarische Archivierungs- und Bereitstellungssystem (BABS) am Münchener Digitalisierungszentrum (MDZ) umgesetzt. Teil des Projektes „BABS“ war die Entwicklung und Umsetzung eines Workflows zur Webarchivierung. Dieser läuft seit 2012 im Regelbetrieb./19/

Innerhalb des Menüpunktes „Literatur“ verfügt die ViFaMusik über eine Sammlung von Internetressourcen (Abb. 1). Diese umfasst mit Stand vom 29.07.2015 bereits 2.979 Websites, von denen 155 langzeitarchiviert sind.

Es gibt zwei Möglichkeiten, die Sammlung von Websites zu durchsuchen. Die erste Suchmöglichkeit besteht in einer inhaltlichen Auswahl verschiedener Kriterien: Thema, Region, Epoche und Publikationstyp. Dabei erfolgt eine Art Browsing durch die Kollektion der Websites. Die zweite Variante ist der klassische Vorgang mit der Eingabe eines Suchbegriffes in einem Suchschlitz. Die Trefferpräsentation ist in beiden Fällen identisch. Wird der Titel einer Website in der Trefferliste angeklickt, gelangt der Nutzer zur Titelaufnahme der ausgewählten Internetressource. Darin sind die Erschließungsinformationen, ein Live-Link und gegebenenfalls der Link zum Webarchiv der BSB enthalten (nur wenn die Website schon langzeitarchiviert ist). Der Button „Online lesen“ in der Trefferliste führt direkt zum Live-Link einer Website. Innerhalb der Titelaufnahme einer Website führt der Link „Webarchiv“ zur WayBack Machine der BSB und bietet damit Zugriff auf die langzeitarchivierte Internetressource mit ihren Webschnitten. Beim Öffnen eines solchen Webschnitts erfolgt der Hinweis darauf, dass es sich um eine archivierte Website handelt, deren Funktionalitäten eventuell



# VIRTUELLE FACHBIBLIOTHEK MUSIKWISSENSCHAFT



## Internetressourcen der ViFaMusik

Suche  Merkliste

Erweiterte Suche

### ▼ Thema

- Aufführungspraxis und Notation (462)
- Formen und Gattungen (355)
- Instrumentenkunde (233)
- Jazz (51)
- Mechanische Musikinstrumente (15)

### ▼ Region

- Afrika (3)
- Amerika und Nordamerika (97)
- Antike Welt (6)
- Asien und Mittlerer Osten (2)
- Australien und Ozeanien (3)
- Deutschland (915)

### ▼ Epoche

- 20. Jahrhundert (704)
- 21. Jahrhundert (593)
- Altertum (13)
- Mittelalter (101)
- Neuzeit bis 1900 (705)

### ▼ Publikationstyp

- Bibliographien, Kataloge, Verzeichnisse (566)
- Einzelne Volltexte und Multimediadateien (gemeinfrei) (501)
- Institutionen (1378)
- Nachschlagewerke (106)
- Periodika, Zeitschriften, usw.

Anzahl: 2973

« 1 2 3 4 5 6 7 8 9 »

1.  **Jahrestagung der Gesellschaft für Musikforschung 2013**  
Die Webseite enthält das Programm der Jahrestagung der Gesellschaft für Musikforschung 2013. [Redaktion ViFa Musik] 2013
  2.  **Sonderforschungsbereich 948 Helden - Heroisierungen - Heroismen (SFG 948)**  
Die Homepage bietet neben einem Kurzprofil des Sonderforschungsgebiets SFB 948 einen Überblick über die Teilbereiche des Projekts ... 2012
  3.  **ACT Zeitschrift für Musik und Performance**  
Die elektronische Zeitschrift Act. Zeitschrift für Musik und Performance ist ein interdisziplinäres Publikationsorgan des firt (Fo... 2012
  4.  **Sound und Performance**  
Die Webseite Sound und Performance steht in enger Verbindung zum 11. Kongress der Gesellschaft für Theaterwissenschaft an der Univ... 2012
  5.  **Katalog der Bibliothek der Hochschule für Musik und Theater Leipzig**  
Der neue Online-Katlog (als Beta-Version) der Hochschule für Musik und Theater "Felix Mendelssohn Bartholdy" Leipzig bei... 2012
  6.  **Musikrecherche**  
Der Aufsatz "Musikrecherche" der Autorin Susanne Hein wurde 2012 vom Deutschen Musikinformationszentrum veröffentlicht. ... 2012
  7.  **Anicius Manlius Severinus Boethius**  
Der spätantike römische Gelehrte Anicius Severinus Boethius (um 480/485-524/526) verfasste Übersetzungen, Kommentare und Lehrbüche... 2012
  8.  **È caso da intermedio! Comic Theory, Comic Style and the Early Intermezzo**  
Dissertation von Keith Johnston über "Comic Theory, Comic Style and the Early Intermezzo" in Italien zwischen 1660 und 1723. [Red... 2012
  9.  **Landesarchiv Baden Württemberg - Internetrundgang Musikalische Fragmente**  
Der Text "Musikalische Fragmente" nimmt Bezug zur gleichnamigen Ausstellung des Landesarchivs Baden-Württemberg im Jahr ... 2011
- Evaluation of Big Noise, Sistema Scotland**

1: Screenshot Sammlung Internetressourcen der ViFaMusik vom 29.07.2015

eingeschränkt sind. Unter dem Menüpunkt „Tutorial“ der ViFaMusik ist ein Video zu finden, in welchem die Suche innerhalb der Internetressourcen erklärt wird.

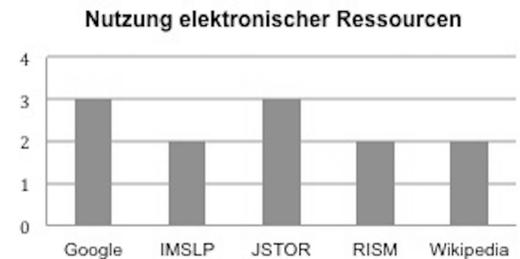
## Forschungsprojekt „Anforderungen aus der Musikwissenschaft an die Erschließung und Archivierung von Internet-Ressourcen“ /20/

Im Verlauf des Antragsprozesses für den Fachinformationsdienst Musikwissenschaft wurde der Nutzen der Webarchivierung für die musikwissenschaftliche Fachcommunity diskutiert. Diese Fragestellung ist im Rahmen eines Forschungsprojektes des berufsbegleitenden Masterstudiengangs Bibliotheks- und Informationsmanagement an der Hochschule der Medien Stuttgart untersucht worden. Die Verfasserin hat das Projekt von März bis August 2014 durchgeführt. Die grundlegenden Fragestellungen setzten sich wie folgt zusammen: Welche Bedürfnisse haben Musikwissenschaftlerinnen und Musikwissenschaftler bezüglich der Archivierung und Erschließung von Internetressourcen? Gibt es einen Nutzen für die musikwissenschaftliche Fachcommunity in Deutschland? Wie gestaltet sich die Quellennutzung von Musikwissenschaftlerinnen und Musikwissenschaftlern in Bezug auf Websites? Welche Kritikpunkte und Wünsche gibt es für die Webarchivierung im Rahmen der ViFaMusik?

Die Durchführung des Projektes basierte auf einer qualitativen Erhebungsmethode: leitfadengestützte Experteninterviews. /21/ Dabei wurden acht Musikwissenschaftlerinnen und Musikwissenschaftler (promoviert oder habilitiert) aus dem deutschsprachigen Raum interviewt. Aufgrund dieser Anzahl ist das Projekt nicht repräsentativ. Es sollte ein erstes Stimmungsbild zur Thematik eingefangen werden.

Alle Expertinnen und Experten verknüpfen etwas mit dem Begriff „Webarchivierung“. Es treten jedoch Abstufungen auf. Zwei Personen hatten nur eine vage Idee und konnten diese nur schwer formulieren. Bei vier Personen war die Vorstellung allgemein gehalten. Zwei der Befragten hatten eine

sehr detaillierte Sicht auf die Webarchivierung und brachten diese von selbst in einen Zusammenhang zum Wert einer Website als wissenschaftliche Quelle. Alle Interviewten gaben übereinstimmend an, dass sie am meisten auf gedruckte Quellen zurückgreifen. Die Nutzung elektronischer Quellen ist jedoch steigend. In vielen Fällen ist dies abhängig vom bearbeiteten Forschungsthema. Websites spielen innerhalb der elektronischen Ressourcen noch eine untergeordnete Rolle. Es dominieren Datenbanken und Suchmaschinen (Abb. 2).



2: Nutzung elektronischer Ressourcen durch die befragten Forschenden  
 Grafik: Barbara Lenk

In diesem Zusammenhang wurden zwei wichtige Aspekte genannt. Websites können Inhalte liefern, die über gedruckte Quellen nicht zu erhalten sind. Andererseits ist die Anzahl an guten fachlichen Websites noch gering. Gleichzeitig wird ein Überangebot an Internetressourcen wahrgenommen. Der einzelne Forschende muss selbst eine Entscheidung bezüglich der Qualität einer Website treffen.

Fünf der befragten Personen kommunizierten einen Bedarf zur Erschließung und Archivierung von Websites. Zwei Interviewte waren skeptisch und wünschten sich mehr Informationen zur Thematik. Ein Experte äußerte sich nicht zu dieser Frage. Die folgenden Anforderungen an die Webarchivierung und Erschließung von Internetressourcen wurden in den Interviews aufgezeigt. Die Dauerhaftigkeit und Nachhaltigkeit bei der Pflege eines Webarchivs spielen eine große Rolle – gerade auch in Verbindung mit der Langzeitarchivierung von Websites. Weitere wichtige Faktoren bilden die Verschlagwortung und Durchsuchbarkeit

multimedialer Inhalte sowie die Erschließung von Websites. Optimal wäre dies in Verbindung mit einer Volltextsuche. Außerdem sollte bei den Internetressourcen eine Qualitätskontrolle erfolgen – bezogen auf die Inhalte und deren Verfügbarkeit.

Sieben von acht Expertinnen und Experten kennen die ViFaMusik und ihre Sammlung von Internetressourcen, nannten dabei aber überwiegend negative Aspekte, z. B. wenig Inhalte, eine fehlende thematische Breite oder mangelnde Transparenz. Gleichzeitig wurden Lösungsvorschläge gemacht: Man wünschte sich eine größere inhaltliche Breite, verbesserte Suchfunktionen und Mechanismen zur Ordnung der Inhalte, Einrichtung von RSS-Feeds oder mehr Aktivitäten zur Vermittlung der ViFaMusik in der Fachcommunity.

Generell ist ein Bedarf zur Webarchivierung klar zu erkennen. Insbesondere erkennen die Personen, welche Websites bereits als wissenschaftliche Quellen nutzen, die Vorteile der Webarchivierung deutlich und sind sich des Problems der Zitierfähigkeit von Internetressourcen bewusst. Die Nutzung elektronischer Ressourcen steigt in der Mu-

sikwissenschaft zwar an, jedoch dominieren noch die gedruckten Quellen. Grund ist ein Mangel an fachlich hochwertigen Websites in Verbindung mit einem Überangebot an Websites ohne qualitative Bewertung. Websites und deren Status als wissenschaftliche Quelle werden in einem Zusammenhang mit einer Generationenfrage gebracht. Die sogenannten „Digital Natives“ gebrauchen elektronische Ressourcen und somit auch Websites ohne Scheu. Die Entwicklung eines Webarchivs kann dabei helfen, die Website als musikwissenschaftliche Quelle besser zu etablieren. Durch langfristige und dauerhafte Archivierung wird eine Website zitierfähig und ihr Wert als wissenschaftliche Quelle geschützt. Die ViFaMusik stellt mit ihrer Kollektion an Internetressourcen ein hilfreiches Angebot dar. Dies ist jedoch nur möglich, wenn die zuvor kritisierten Sachverhalte ausgemerzt werden.

Barbara Lenk leitete bis Juni 2015 die Bibliothek der Nürnberger Hochschule für Musik und ist seit Juli 2015 die Bibliotheksleiterin der Hochschule für Bildende Künste in Dresden.

- 1 Der vorliegende Artikel basiert auf einem Vortrag, den die Verfasserin am 24.06.2015 auf der IAML/IMS-Tagung 2015 in New York City gehalten hat.
- 2 Der eigentliche Ursprung ist das Advanced Research Projects Agency Network (ARPANET), welches rein militärischer Natur war.
- 3 „Website-Archivierung an der BSB“, [www.babs-muenchen.de/index.html?c=workflows\\_web&tl=de](http://www.babs-muenchen.de/index.html?c=workflows_web&tl=de) (23.07.2015).
- 4 Vgl. Adrian Brown: *Archiving websites*, London 2006, S. 3, und „Website-Archivierung an der BSB“.
- 5 Vgl. Brown: *Archiving websites*, S. 5.
- 6 Vgl. Andreas Rauber, Hans Liegmann: Webarchivierung zur Langzeiterhaltung von Internet-Dokumenten, in: *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, hrsg. von Heike Neuroth, Achim Obwald, Regine Scheffel u. a., Version 2.3 – 2010, S. 92, und „Website-Archivierung an der BSB“.
- 7 Die WayBack Machine ist eine Entwicklung des Internet Archives und ermöglicht es, die Webschnitte einer Website anhand eines Zeitstrahls in chronologischer Auflistung zu verfolgen und darauf zuzugreifen. Siehe hierfür: <https://archive.org/index.php> (28.07.2015).
- 8 Vgl. Rauber, Liegmann: Webarchivierung, S. 98, 100.

- 9 Vgl. „Website-Archivierung an der BSB“ und Rauber, Liegmann: Webarchivierung, S. 93.
- 10 Vgl. Rauber, Liegmann: Webarchivierung, S. 97.
- 11 Ebd., S. 94.
- 12 Vgl. Brown, *Archiving websites*, S. 3.
- 13 Vgl. Rauber, Liegmann: Webarchivierung, S. 97.
- 14 Vgl. „Website-Archivierung an der BSB“ und Rauber, Liegmann: Webarchivierung, S. 98.
- 15 Vgl. Rauber, Liegmann: Webarchivierung, S. 98.
- 16 Zeitgenössische Komponistinnen und Komponisten nutzen ihre Websites in vielen Fällen zum Vertrieb der eigenen Noten.
- 17 Siehe: [www.vifamusik.de/startseite.html](http://www.vifamusik.de/startseite.html) (29.07.2015).
- 18 Die ViFaMusik war vorher ebenfalls ein Teil des Angebotes im Sondersammelgebiet Musikwissenschaft.
- 19 Vgl. „Website-Archivierung an der BSB“.
- 20 Das gesamte Projekt kann im zugehörigen Projektbericht nachvollzogen werden: <http://nbn-resolving.de/urn:nbn:de:bsz:900-opus-45384> (30.07.2015).
- 21 Vgl. hierzu Jochen Gläser, Grit Laudel: *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*, 4. Aufl., Wiesbaden 2010, und Udo Kuckartz: *Qualitative Inhaltsanalyse*, 2., durchgeseh. Aufl., Weinheim 2014.