

Alexis Rossi:  
**Internet Archive**

aus:

**Föderale Vielfalt - Globale Vernetzung.**

Strategien der Bundesländer Strategien der Bundesländer für das kulturelle Erbe in der digitalen Welt.

Herausgegeben von Ellen Euler und Paul Klimpel

Schriftenreihe:

»Kulturelle Erbe in der digitalen Welt«. Band 2

S. 224–237



Hamburg University Press

Verlag der  
Staats- und Universitätsbibliothek Hamburg  
Carl von Ossietzky



Eine Publikation der  
Deutschen Digitalen Bibliothek



---

## Impressum

**Band 2** der Schriftenreihe »Kulturelles Erbe in der digitalen Welt«  
verantwortet von Ellen Euler  
1. Auflage, 2016

Verlag: Hamburg University Press, Verlag der Staats- und Universitätsbibliothek  
Hamburg Carl von Ossietzky, Von-Melle-Park 3, 20146 Hamburg, [hup.sub.uni-hamburg.de](http://hup.sub.uni-hamburg.de)  
Permalink: [hup.sub.uni-hamburg.de/purl/HamburgUP\\_DDB2\\_Vielfalt](http://hup.sub.uni-hamburg.de/purl/HamburgUP_DDB2_Vielfalt)

Herausgeber: Stiftung Preußischer Kulturbesitz, Von-der-Heydt-Straße 16–18, 10785 Berlin,  
gesetzlich vertreten durch ihren Präsidenten Herrn Prof. Dr. Dr. h.c. mult. Hermann Parzinger,  
handelnd für das durch Verwaltungs- und Finanzabkommen zwischen Bund und Ländern  
errichtete „Kompetenznetzwerk Deutsche Digitale Bibliothek“.

Projektleitung: Ellen Euler  
Redaktion: Ellen Euler (EE), Paul Klimpel (PK)  
Layout: Beate Stangl / [beworx.de](http://beworx.de)  
Covergestaltung: Kerstin Stäblein / IIIIgrafikstäbleinIII  
Lektorat, Satz und Herstellung: Hofmeister Stauder. Büchermacher, Berlin  
Druck und Weiterverarbeitung: hansadruk und Verlags-GmbH & Co KG, Kiel

Bildstrecke  
Projektleitung: Astrid B. Müller  
Alle Fotos: Deutsche Digitale Bibliothek, Fotograf: Jürgen Keiper (Alle Rechte vorbehalten)

Texte und Fotos stehen, soweit nicht anders gekennzeichnet, unter der Creative-Commons-  
Lizenz Namensnennung – Keine Bearbeitungen 4.0 (CC BY ND 4.0). Das bedeutet, dass sie  
vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden dürfen, auch kommerziell,  
sofern dabei stets die Urheber, die Quelle des Textes und o.g. Lizenz genannt wird, deren  
genaue Formulierung Sie unter <https://creativecommons.org/licenses/by-nd/4.0/legalcode>  
nachlesen sollten.

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen  
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de>  
abrufbar.

### **Deutsche Digitale Bibliothek, Geschäftsstelle**

Von-der-Heydt-Straße 16–18, 10785 Berlin, Telefon +49 30 266 411 432  
[geschaeftsstelle@deutsche-digitale-bibliothek.de](mailto:geschaeftsstelle@deutsche-digitale-bibliothek.de),  
[www.deutsche-digitale-bibliothek.de](http://www.deutsche-digitale-bibliothek.de)

Die Deutsche Digitale Bibliothek ist ein Kooperationsprojekt von Bund, Ländern und  
Kommunen und wird gefördert durch die Beauftragte der Bundesregierung für Kultur und  
Medien aufgrund Beschlusses des Deutschen Bundestags sowie der 16 Bundesländer.

ISSN (PRINT) 2509-8276                      ISSN (ONLINE) 2509-8284  
ISBN (PRINT) 978-3-943423-34-1

Elektronische Ausgaben sind in den folgenden Formaten frei verfügbar:  
ISBN (EPUB) 978-3-943423-35-8            ISBN (PDF) 978-3-943423-36-5

# Inhalt

## **Föderale Vielfalt – Globale Vernetzung**

Ellen Euler und Paul Klimpel ..... 8

## **Föderale Vielfalt**

### **Baden-Württemberg**

Ursula Bernhardt für das Ministerium für  
Wissenschaft, Forschung und Kunst Baden-Württemberg ..... 16

### **Bayern**

Bayerisches Staatsministerium für  
Bildung und Kultus, Wissenschaft und Kunst ..... 26

### **Berlin**

Anja Müller und Beate Rusch für die  
Senatskanzlei – Kulturelle Angelegenheiten Berlin ..... 38

### **Brandenburg**

Ulf Preuß für die Koordinierungsstelle Brandenburg-digital ..... 52

### **Bremen**

Anna Greve für den Senator für  
Kultur der Freien Hansestadt Bremen ..... 72

### **Hamburg**

Anke Hönnig und Horst Scholz für die  
Kulturbehörde der Freien und Hansestadt Hamburg ..... 88

### **Niedersachsen**

Till Manning für das Niedersächsische  
Ministerium für Wissenschaft und Kultur ..... 102

### **Nordrhein-Westfalen**

Frank Michael Bischoff, Beate Möllers, Irmgard Siebert, Beate  
Tröger, Renate Vogt für das Ministerium für Familie, Kinder, Jugend,  
Kultur und Sport des Landes Nordrhein-Westfalen ..... 112

### **Sachsen**

Frank Aurich für das Sächsische Staatsministerium  
für Wissenschaft und Kunst ..... 126

## **Schleswig-Holstein**

Martin Lätzel für das Ministerium für Justiz,  
Kultur und Europa des Landes Schleswig-Holstein ..... 136

## **Thüringen**

Carsten Pettig für die Thüringer Staatskanzlei ..... 152

## **Die digitalisierte Gesellschaft und kommunale Kultureinrichtungen**

Raimund Bartella für den Deutschen Städtetag ..... 166

## **Globale Vernetzung**

### **„Digitales Archiv“ und „Kulturelles Gedächtnis“ in Japan – unter besonderer Berücksichtigung der Initiativen der Nationalen Parlamentsbibliothek Japans**

Toshiyasu Oba ..... 184

### **Die Raison d'Être nationaler Bibliotheken im digitalen Zeitalter. Die norwegische Nationalbibli- othek auf der Suche nach neuer Bedeutung**

Roger Jøsevoid ..... 206

## **Internet Archive**

Alexis Rossi ..... 224

# Internet Archive

Alexis Rossi

Das Internet Archive hat den Auftrag, für einen „universellen Zugang zum gesamten Wissen“ zu sorgen. Es ist unser Ziel, eine öffentliche Bibliothek für die gesamte Weltöffentlichkeit zu erstellen. Dies erfordert das Sammeln von Informationen, deren sichere Speicherung, den Aufbau einer Infrastruktur, die Millionen von Menschen versorgt, und die Beschäftigung mit den Themen Kundendatenschutz und Informationszugang.

Die Organisation wurde im Jahr 1996 von Brewster Kahle gegründet. Sie ist gemeinnützig und keiner staatlichen Einrichtung angegliedert. Internet Archive bedient zwei bis drei Millionen Menschen pro Tag.

Wir arbeiten mit vielen Partnern zusammen, darunter Bibliotheken, Museen, Archiven und Einzelpersonen, um unsere öffentlichen Sammlungen aufzubauen. Jeder kann kostenlos Medien auf [archive.org](https://archive.org) hochladen, um sie für die Zukunft zu bewahren, und wir empfehlen den Beitragenden, ihre Einträge mit Creative-Commons-Lizenzen zu versehen. Die Finanzierung der Organisation stammt aus vielen Quellen, unter anderem aus der Digitalisierung von Büchern, der Web-Archivierung, aus Spenden, Beihilfen und aus Stiftungen.

## Archivierung von Webseiten

Das Web ist kurzlebig. Die durchschnittliche Lebensdauer einer Webseite beträgt 100 Tage, bevor sie sich ändert oder verschwindet. Während das Web 1996 eine weniger integrale Rolle in unserem täglichen Leben spielte als heute, stellte sich Brewster Kahle vor, dass es eines Tages unser Leben genauso umfassend widerspiegeln könnte wie die Tageszeitung. Wir müssen uns an unsere Vergangenheit erinnern, um aus ihr zu lernen, und das Speichern von Webseiten schien ein wichtiger Schritt zu sein, um eine globale Bibliothek aufzubauen.

Die erste öffentliche Sammlung von Internet Archive wurde 1997 gemeinsam mit der Smithsonian Institution<sup>1</sup> veröffentlicht und enthielt eine Auswahl an Webseiten zu den amerikanischen Präsidentschaftswahlen 1996. 2000<sup>2</sup> haben wir unsere Bemühungen mit einer Pilotstudie zur Webarchivierung gemeinsam mit der amerikanischen Library of Congress ausgeweitet.

Die Wayback Machine<sup>3</sup>, in die man eine URL zur Ansicht archivierter Versionen von Webseiten eingeben kann, wurde 2001<sup>4</sup> eingeführt. Die Sammlung umfasste zu der Zeit etwa zehn Milliarden Webressourcen. Vor der Einführung war dieses Archiv nur Forschern zugänglich, die zudem über fortgeschrittene technische Fähigkeiten verfügen mussten, um es zu durchsuchen. Die Notwendigkeit für ein Internet-Archiv wurde zu der Zeit von einigen Personen infrage gestellt, 20 Jahre später enthält die Wayback Machine jedoch die einzige öffentliche Aufzeichnung über die Anfangszeit dieses wichtigen Kommunikationsmediums.

Der Archive-It<sup>5</sup>-Service wurde von uns 2005<sup>6</sup> eingeführt, um andere Institutionen bei der Sicherung von Webressourcen zu unterstützen, die sie für wichtig erachteten. Archive-It.org stellt Tools zur Verfügung, die es technisch nicht so versierten Personen ermöglichen, Webressourcen zu crawlen und zu archivieren. Der Dienst ist im Laufe der Jahre gewachsen und hat heute mehr als 400 Partner, die mehrere Tausend gut kuratierte Sammlungen erstellt haben. Wir arbeiten außerdem mit mehreren Nationalbibliotheken aus der ganzen Welt

1 Vgl. <http://web.archive.org/web/19970126045828/http://www.archive.org/> (Letzter Aufruf: 20.05.2016).

2 Vgl. <http://web.archive.org/web/20010203130300/http://archive.org/news/index.html#2000LOC> (Letzter Aufruf: 20.05.2016).

3 Vgl. <http://archive.org/web/> (Letzter Aufruf: 20.05.2016).

4 Vgl. <http://web.archive.org/web/20011130142035/http://archive.org/> (Letzter Aufruf: 20.05.2016).

5 Vgl. <http://archive-it.org> (Letzter Aufruf: 20.05.2016).

6 Vgl. <http://web.archive.org/web/20051124234136/http://www.archive-it.org/> (Letzter Aufruf: 20.05.2016).

zusammen, um umfangreiche Crawls ihrer nationalen Domänen zu vervollständigen.

Während das Internet an Größe und Wichtigkeit zugelegt hat, haben wir unser eigenes Archiv an Webressourcen ebenfalls vergrößert. Internet Archive erfasst derzeit etwa eine Milliarde Web-Captures pro Woche. Die Wayback Machine ist auf etwa 470 Milliarden Ressourcen angewachsen.

Jeder kann eine URL in die Wayback Machine eingeben und alte Versionen von Webseiten einsehen. Die Wayback Availability API<sup>7</sup> ermöglicht es Entwicklern, archivierte Ressourcen aufzufinden, was derzeit für die Reparatur defekter Links auf Webseiten verwendet wird. Außerdem erforschen wir Wege, die das Durchsuchen dieses gewaltigen Werks durch die Stichwortsuche vereinfachen.

## Fernseharchiv

Internet Archive hat Ende 2000 damit begonnen, Fernsehsendungen zu archivieren, zunächst von 20 Sendern aus den USA und mehreren anderen Ländern. Zu der Zeit war Online-Video-Streaming unüblich und wir waren uns nicht sicher, wie wir Zugriff auf diese archivierten Inhalte bieten können. Aber das Fernsehen ist im Wesentlichen ein kurzlebiges Medium, ebenso wie das Web – es muss direkt während der Entstehung konserviert werden, andernfalls ist die Information unter Umständen für immer verloren. Während einige Unterhaltungsprogramme regelmäßig wiederholt werden, wird aktuelleres Material wie politische Debatten oder Nachrichtensendungen ein- oder zweimal wiederholt und verschwindet dann aus dem öffentlichen Blickfeld.

Die Terroranschläge in den Vereinigten Staaten am 11. September 2001 waren ein globales Ereignis, welches rund um die Welt im Fernsehen gesehen und verbreitet wurde. Dies hat das Internet Archive dazu veranlasst, einen kleinen Teil des archivierten Materials zur Verfügung zu stellen. Am 11. Oktober 2001<sup>8</sup> wurde das Nachrichtenprogramm einer Woche rund um den 11. September veröffentlicht. Diese Sammlung sollte kein Denkmal errichten; sie sollte als Werkzeug für Forscher/innen und Historiker/innen dienen. Nachrichtenmeldungen im Fernsehen beeinflussen Menschen und Ereignisse. Wir können

<sup>7</sup> Vgl. [http://archive.org/help/wayback\\_api.php](http://archive.org/help/wayback_api.php) (Letzter Aufruf: 20.05.2016).

<sup>8</sup> Vgl. <https://web.archive.org/web/20011013040507/http://tvnews3.televisionarchive.org/tvarchive/html/index.html> (Letzter Aufruf: 20.05.2016).



nicht unsere eigene Geschichte erforschen und wiedergeben, wenn wir keinen Zugriff auf sie haben.

Wir haben unsere Fähigkeiten bei der Fernseharchivierung 2009 erweitert und damit begonnen, 60 Sender aus den USA und anderen Ländern aufzuzeichnen. 2012 starteten wir das Television News Archive<sup>9</sup>. Mit diesem Dienst kann man die Untertitel ausgewählter amerikanischer Nachrichtensendungen zwischen 2009 und gestern durchsuchen; wir warten 24 Stunden, bevor wir dem Dienst neue Sendungen hinzufügen. Die Seite beinhaltet derzeit mehr als eine Million durchsuchbare Sendungen. Die Nutzer finden Suchtreffer für ihre Schlüsselwörter und können kurze Ausschnitte des Videos zu ihren Ergebnissen abspielen. Diese Ausschnitte können geteilt und zitiert werden. Um mehr als nur diese kurzen Schnipsel zu sehen, können die Nutzer eine DVD mit dem vollständigen Beitrag ausleihen.

Für die Forscher/innen erwiesen sich die Medien im Television News Archive als nützlich. Wissenschaftler/innen nutzten unsere hauseigenen virtuellen Maschinen, um Schlagzeilen dahingehend zu analysieren, wie oft geografische Orte erwähnt<sup>10</sup> und über welche Präsidentschaftskandidaten in den Nachrichtensendungen diskutiert werden.<sup>11</sup> Die Sammlung wurde außerdem dazu genutzt, ein an der Columbia University entwickeltes Werkzeug für das Audio-Fingerprinting mit der Bezeichnung *audfprint*<sup>12</sup> zu verbessern. Diese Technologie hilft bei der Nachverfolgung, wie sich Videoclips durch die Nachrichtensendungen ausbreiten. Beispielsweise offenbart die Analyse einer politischen Debatte, welche „Soundbites“ von Nachrichtensendungen aufgegriffen und für die Zuschauer verstärkt werden.<sup>13</sup> Software wie diese unterstützt uns auf eine neue Weise, die gesammelten archivierten Fernsehsendungen zu erforschen.

Das Political Ad Archive<sup>14</sup> wurde im Januar 2016 gestartet.<sup>15</sup> Es verfolgt amerikanische politische TV-Werbespots, die in Vorwahl-Staaten gesendet werden, und bringt diese mit Informationen über Finanzierungsquellen und mit durch Journalisten überprüfte Fakten in Verbindung. Metadaten zu die-

9 Vgl. <https://archive.org/details/tv> (Letzter Aufruf: 20.05.2016).

10 Vgl. [www.theatlantic.com/technology/archive/2013/12/a-new-map-reveals-the-geography-of-american-tv-news/282443/](http://www.theatlantic.com/technology/archive/2013/12/a-new-map-reveals-the-geography-of-american-tv-news/282443/) (Letzter Aufruf: 20.05.2016).

11 Vgl. [www.theatlantic.com/politics/archive/2015/08/graphic-whos-the-most-popular-candidate-mentioned-on-television/402451/](http://www.theatlantic.com/politics/archive/2015/08/graphic-whos-the-most-popular-candidate-mentioned-on-television/402451/) (Letzter Aufruf: 20.05.2016).

12 Vgl. <http://labrosa.ee.columbia.edu/matlab/audfprint/> (Letzter Aufruf: 20.05.2016).

13 Vgl. [http://television.gdeltproject.org/cgi-bin/iatv\\_campaign2016\\_rdebate1/iatv\\_campaign2016\\_rdebate1\\_prime](http://television.gdeltproject.org/cgi-bin/iatv_campaign2016_rdebate1/iatv_campaign2016_rdebate1_prime) (Letzter Aufruf: 20.05.2016).

14 Vgl. <http://politicaladarchive.org> (Letzter Aufruf: 20.05.2016).

15 Vgl. [www.pbs.org/newshour/rundown/a-new-free-tool-thats-like-x-ray-glasses-for-political-ads/](http://www.pbs.org/newshour/rundown/a-new-free-tool-thats-like-x-ray-glasses-for-political-ads/) (Letzter Aufruf: 20.05.2016).

sen Werbespots, einschließlich Geldgebern, Fernsehmärkten und Häufigkeit, wie oft Werbespots gesendet wurden, können heruntergeladen werden. Diese Ressource wurde von Journalisten zur Veröffentlichung vieler Artikel über die Wahlen 2016<sup>16</sup> genutzt.

## Videoarchiv

Unser erstes Videoarchiv wurde 2001 veröffentlicht.<sup>17</sup> Wir haben mit Rick Prelinger, einem Filmsammler und Historiker, zusammengearbeitet, um digitale Kopien von etwa 1000 Filmen, die nicht im Kino gezeigt worden sind, aus seinem Archiv zu bewahren. 2001 waren wir nicht in der Lage, diese Videos online zu streamen. Um sie zu sehen, mussten die Nutzer sie – über sehr langsame Verbindungen – herunterladen. Die Filme haben aber dennoch ihre Fans gefunden und die Sammlung wuchs im Laufe der Jahre stetig. Während die meisten der Videos heruntergeladen werden können, ist bei anderen Medien nur das Online-Streaming erlaubt.

Heute gibt es mehr als zwei Millionen Videos auf archive.org und die Prelinger-Sammlung wurde durch Spielfilme in Gesamtlänge, Live-Vorstellungen, Cartoons, Lesungen, Video-Blogs, Nachrichtenberichte, Dokumentationen und andere Genres ergänzt.

## Audioarchiv

Die erste Sammlung von Audiodateien auf archive.org wurde 2002 veröffentlicht.<sup>18</sup> Unter dem Namen Etree<sup>19</sup> trafen sich Freiwillige aus der ganzen Welt, die Live-Konzerte aufnahmen und die Musik dann miteinander teilten. Die Gruppe hat ausschließlich Künstler aufgenommen, deren Aufnahme Richtlinien nicht so streng waren. Dies ist eine Tradition, die mit den *Grateful Dead* begann – Fans zeichneten Konzerte auf und vertrieben dann die Audiokassetten

<sup>16</sup> Vgl. <http://politicalarchive.org/press/> (Letzter Aufruf: 20.05.2016).

<sup>17</sup> Vgl. <http://web.archive.org/web/20010331221954/http://archive.org/> (Letzter Aufruf: 20.05.2016).

<sup>18</sup> Vgl. <http://web.archive.org/web/200209/http://www.archive.org/> (Letzter Aufruf: 20.05.2016).

<sup>19</sup> Vgl. <http://etree.org>

untereinander. Etree hatte diese Tradition ins Internet gebracht, aber ihr waren durch Speicherkapazität und Bandbreite Grenzen gesetzt. Sie konnte ihrer Fangemeinde nur wenige Konzerte auf einmal anbieten.

Internet Archive ist auf Etree zugegangen und bot der Gruppe an, alle aufgezeichneten Shows zu speichern und zu jeder Zeit kostenlos zugänglich zu machen. Sie stimmten einem Versuch zu und bis heute hat diese eine Gruppe Freiwilliger über 150 000 Live-Konzerte von mehr 6000 Bands beigesteuert.<sup>20</sup>

Eine weitere Gruppe bildete sich zum Thema Hörbuch. Librivox.org ist eine Gemeinschaft Freiwilliger, die Lesungen gemeinfreier Bücher gestaltet. Sie haben im Laufe der Jahre mehrere Tausend Bücher aufgenommen und viele dieser Bücher wurden millionenfach von archive.org heruntergeladen.<sup>21</sup>

Weitere Sammlungen von Audiodateien folgten kurz darauf und heute umfasst die Seite mehr als zwei Millionen Audiodateien, unter anderem alte Hörfunksendungen, Podcasts, religiöse Predigten, alte 78er-Schallplatten, moderne Radiobeiträge und vieles mehr. Während die meisten dieser Artikel heruntergeladen werden können, können einige von ihnen auf Wunsch der Urheber nur gestreamt werden.

## Textarchiv

Unsere ersten E-Book-Sammlungen stellte Internet Archive 2002 online.<sup>22</sup> Um zu zeigen, was mit ihnen möglich ist, haben wir einen digitalen Bücherbus<sup>23</sup> entwickelt – einen Kleinbus mit einer Satellitenschüssel für das Internet und einer Ausrüstung, mit der man die gemeinfreien elektronischen Bücher aus unserem Archiv drucken und binden kann. Die Produktion dieser Bücher ist so preiswert, dass es einfacher ist, eine Kopie des Buchs zu erstellen und wegzugeben, anstatt es zu verleihen. Dieser Bücherbus wurde in Indien und Ägypten nachgemacht, um den Menschen, die keinen Zugang zu Bibliotheken haben, Bücher nahezubringen.

Als 2005 das Projekt Google Books startete, begann Internet Archive mit der Entwicklung eines eigenen Digitalisierungsprogramms. Wir hatten Beden-

<sup>20</sup> Vgl. <http://archive.org/details/etree> (Letzter Aufruf: 20.05.2016).

<sup>21</sup> Vgl. <https://archive.org/details/librivoxaudio> (Letzter Aufruf: 20.05.2016).

<sup>22</sup> Vgl. <http://web.archive.org/web/20021001124437/http://www.archive.org/> (Letzter Aufruf: 20.05.2016).

<sup>23</sup> Vgl. [http://web.archive.org/web/20021010095421/http://webdev.archive.org/texts/bookmobile-open\\_house.php](http://web.archive.org/web/20021010095421/http://webdev.archive.org/texts/bookmobile-open_house.php) (Letzter Aufruf: 20.05.2016).

ken, dass Bibliotheken unter Umständen Bücher bei Google Books digitalisieren und dann das analoge Buch unzugänglich machen könnten, wodurch nur eine kommerzielle Organisation über das Wissen verfügen würde. Eine Reihe von Partnern hat sich uns in der Open Content Alliance angeschlossen, um eine Alternative zu schaffen.<sup>24</sup>

Nachdem wir viele handelsübliche Geräte zur Bücherdigitalisierung getestet haben, stellten wir fest, dass wir unser eigenes zerstörungsfreies Scansystem entwickeln müssen. Wir entwarfen ein Gerät mit dem Namen „Scribe“, mit dem der Anwender qualitativ sehr hochwertige, gut belichtete Fotos von Buchseiten machen kann. Wir entwickelten außerdem eine Software, um die daraus entstandenen Bilder in ansprechende E-Books umzuwandeln, einschließlich der Begradigung der Textzeilen und des Zurechtschneidens der Seiten. Wir verwenden kommerzielle optische Zeichenerkennungssoftware (optical character recognition – OCR), um die Bilder der Buchseiten in Text umzuwandeln, EPUB-Dateien zu produzieren und um es den Nutzern zu ermöglichen, die Bücher zu durchsuchen. Wir haben kürzlich eine portablere Version dieses originalen Geräts zur Buchdigitalisierung entwickelt, welches wir als „Table Top Scribe“ bezeichnen.<sup>25</sup>

Unsere ersten Scanprojekte konzentrierten sich auf die Digitalisierung älterer Bücher. Diese sind in Bibliotheken eher selten im Umlauf und es besteht die Gefahr, dass sie eingelagert oder ausrangiert werden. Diese digitalisierten älteren Bücher stehen in vielen Formaten zum Download zur Verfügung. Es können auch Bücher online unter Verwendung der von uns entwickelten Open-Source-Software „Bookreader“<sup>26</sup> eingesehen werden.

Es war vorgesehen, dass die Open Library<sup>27</sup>, die 2008 eingeführt wurde,<sup>28</sup> für jedes jemals herausgegebene Buch je eine eigene Webseite hat. Wir haben Aufzeichnungen von Metadaten aus vielen Quellen zusammengestellt und diese nach Möglichkeit mit den elektronischen Versionen der Bücher verlinkt. OpenLibrary.org enthält heute Aufzeichnungen von mehr als 20 Millionen Büchern, auf die 250 000 Nutzer pro Tag zugreifen.

24 Vgl. <http://web.archive.org/web/20051007010920/http://www.opencontentalliance.org/> (Letzter Aufruf: 20.05.2016).

25 Vgl. <http://blog.archive.org/2015/10/22/special-book-collections-come-online-with-the-table-top-scribe/> (Letzter Aufruf: 20.05.2016).

26 Vgl. <https://github.com/openlibrary/bookreader> (Letzter Aufruf: 20.05.2016).

27 Vgl. <https://openlibrary.org> (Letzter Aufruf: 20.05.2016).

28 Vgl. <http://web.archive.org/web/20080505074433/http://www.openlibrary.org/> (Letzter Aufruf: 20.05.2016).

Später begannen wir dann mit der Digitalisierung zeitgenössischer Bücher. 2010 fingen wir an, sehbehinderten Menschen diese zeitgenössischen Bücher unter Verwendung von verschlüsselten DAISY-Dateien auf Open Library zur Verfügung zu stellen.<sup>29</sup> Die Nutzer müssen sich beim amerikanischen National Library Service registrieren, um einen Entschlüsselungsschlüssel zu erhalten, mit dem sie die Bücher anhören können.

Anschließend entwickelten wir ein Leihprogramm, um die Bücher einem größeren Nutzerkreis zur Verfügung zu stellen.<sup>30</sup> Wenn wir ein zeitgenössisches Buch digitalisieren, legen wir das Buch in unser reales Archiv und stellen die digitale Kopie zur Ausleihe durch jeweils eine Person zur Verfügung. Die Nutzer können bis zu fünf Bücher auf einmal für einen Zeitraum von zwei Wochen ausleihen. Wenn der Nutzer das Buch online lesen möchte, kann er es durch unsere Bookreader-Software ausleihen. Möchte er eine Kopie herunterladen, muss er die Adobe-Digital-Editions-Software verwenden, die die Dateien schützt.

Das Digitalisierungsprogramm für Bücher ist erheblich gewachsen. Wir haben im Laufe der Jahre mit vielen Bibliotheken zusammengearbeitet, um digitale Kopien ihrer Texte anzufertigen. Derzeit digitalisieren wir etwa 1000 Bücher pro Tag in 30 Scan-Zentren auf fünf Kontinenten. Die Textsammlungen auf archive.org umfassen etwa vier Millionen Werke, zuzüglich weiterer vier Millionen Texte wie Zeitschriftenartikel, Regierungsdokumente, Dokumente über Gerichtsverfahren und Volkszählungslisten.

## Bildarchiv

2008 ging das Internet Archive eine Partnerschaft mit der NASA ein, um digitale Bilder aus vielen Zentren zu sammeln und sie in einer Sammlung zusammenzufassen. Video-, Audio- und Textdateien wurden ebenfalls mit einbezogen und es wurde ein Portal speziell für diesen Zweck eingerichtet.<sup>31</sup> Wir starteten das NASA-Bildarchiv mit etwa 100 000 Bildern.<sup>32</sup>

<sup>29</sup> Vgl. <http://blog.archive.org/2010/11/26/3424/> (Letzter Aufruf: 20.05.2016).

<sup>30</sup> Vgl. <http://blog.archive.org/2011/06/25/in-library-ebook-lending-program-expands-to-1000-libraries/> (Letzter Aufruf: 20.05.2016).

<sup>31</sup> Vgl. <http://web.archive.org/web/20080725062000/http://nasaimages.org/> (Letzter Aufruf: 20.05.2016).

<sup>32</sup> Vgl. <https://archive.org/details/nasa> (Letzter Aufruf: 20.05.2016).

Im Laufe der Jahre haben wir außerdem Kartensammlungen und Bilder von Kunstwerken aus Museen erhalten. Heute umfasst das Bildarchiv mehr als eine Million Objekte.<sup>33</sup>

## Softwarearchiv

Die erste Softwaresammlung erschien 2002 auf archive.org<sup>34</sup>. Wir haben diese Sammlungen im Laufe der Jahre ergänzt, allerdings war es nicht einfach, die Dateien zu nutzen oder auf sie zuzugreifen. Software wird programmiert, damit sie in einer bestimmten Umgebung funktioniert; kann man diese Umgebung nicht nachbilden, dann kann man die Funktionsweise der Software nicht nachempfinden. Mit anderen Worten: Man kann zwar ein Spiel-Modul einlegen, das für eine bestimmte, in den 1980er-Jahren gebaute Spielkonsole entwickelt worden ist, aber man kann das Spiel nicht spielen.

Im Jahr 2010 hat Internet Archive seinen ersten speziell für die Software-Archivierung vorgesehenen Mitarbeiter eingestellt. Dieser Kurator, Jason Scott, half dabei, eine Sammlung für archivierte Software und zugehörige Medien aufzubauen, er ging aber auch das Thema Zugang an. Eine Gruppe Freiwilliger war damit beschäftigt, ein System mit der Bezeichnung JSMESS<sup>35</sup> zu entwickeln, das die Emulation alter Betriebssysteme in einem Web-Browser ermöglicht. Scott hat mit den Freiwilligen zusammengearbeitet, um diese Technik auf der archive.org-Seite einzuführen.

Wir starteten 2013<sup>36</sup> die erste experimentelle Sammlung historischer Software, die man im Browser durch Emulation nachempfinden konnte. Im Laufe der Jahre hat sich die Software stetig verbessert, es wurden weitere Betriebssysteme unterstützt und heute<sup>37</sup> können Millionen von Menschen auf Internet Archive auf historische Software zugreifen, die jahrzehntelang unzugänglich war.

33 Vgl. <https://archive.org/details/image> (Letzter Aufruf: 20.05.2016).

34 Vgl. <http://web.archive.org/web/20021001124437/http://www.archive.org/> (Letzter Aufruf: 20.05.2016).

35 Vgl. <https://github.com/jsmess> (Letzter Aufruf: 20.05.2016).

36 Vgl. <http://blog.archive.org/2013/10/25/microcomputer-software-lives-again-this-time-in-your-browser/> (Letzter Aufruf: 20.05.2016).

37 Vgl. <https://archive.org/details/software> (Letzter Aufruf: 20.05.2016).

## Analoges Archiv

Da analoge Objekte in digitale Formate umgewandelt werden, müssen wir eine Entscheidung darüber treffen, was mit diesen analogen Medien geschehen soll. Es gibt gute Gründe dafür, sie zu erhalten. Sie sind die originalen, maßgeblichen Objekte; wenn es Fragen oder Probleme mit einer digitalen Version gibt, kann das analoge Objekt in der Zukunft als Referenz herangezogen werden. Es dient darüber hinaus als Reserve, falls die digitalen Kopien beschädigt werden.

Unser 2011 gegründetes<sup>38</sup> analoges Archiv hat es uns ermöglicht, mehrere Millionen Bücher, Filme, LPs, CDs, Software, Filme und andere analoge Medien zu erhalten. Wir versuchen, ein Exemplar jedes einzelnen Mediums, das wir beschaffen können, zu bewahren. Dann digitalisieren wir das Objekt, verwenden die digitale Kopie für den Zugriff und lagern das analoge Medium ein. Das analoge Archiv wurde für den langfristigen Erhalt von Materialien eingerichtet, nicht für den täglichen Zugriff; der tägliche Zugriff erfolgt über die digitalen Kopien. Da wir keinen Zugriff auf die analoge Sammlung anbieten, können wir diese Medien sehr kompakt und daher zu niedrigeren Kosten lagern, als dies in vielen Bibliothekslagerräumen möglich ist.

## Erhalt digitaler Daten

Das Internet Archive umfasst derzeit 25 Petabytes an unikalen Daten. Wir speichern alle Daten auf unseren eigenen Servern und es gibt mindestens zwei Sicherungskopien an verschiedenen real existierenden Orten. Wir betreiben Datenzentren in unserem Hauptgebäude in San Francisco und in unseren Archivgebäuden in Richmond, Kalifornien. Zusätzlich liegen Teil-Kopien von Archivdaten in Amsterdam und in der Bibliothek von Alexandria in Ägypten.

Die Speicherung dieser Datenmenge stellt uns vor große Herausforderungen. Wir verfügen über zigtausend Festplatten, sodass ständig Laufwerke versagen, die dann schnell ausgetauscht werden müssen. Wir kontrollieren diese Dateien, um sicherzustellen, dass kein Datenverlust auftritt. Die Kosten für Elektrizität und Bandbreite müssen gezahlt werden.

Aber die größte Herausforderung liegt darin, Medien über längere Zeiträume für die Öffentlichkeit zugänglich zu halten, da die Zugriffsmethoden sich än-

<sup>38</sup> Vgl. <http://blog.archive.org/2011/06/06/why-preserve-books-the-new-physical-archive-of-the-internet-archive/> (Letzter Aufruf: 20.05.2016).

dern. Wenn neue Browser, Tablets oder Telefone auf den Markt kommen, können Dateiformate schnell veralten. Als beispielsweise das erste iPhone auf den Markt kam, waren die Videoveritionen, die wir für den Click-and-Play-Zugriff auf der Seite verwendet haben, nicht kompatibel; plötzlich konnte ein Großteil unserer Nutzer die Videos aus unseren Sammlungen nicht abspielen. Wir haben die originalen Videodateien verwendet, um mehrere Millionen Zugriffsversionen in ein neues Format umzuwandeln. Die Erhaltung der Zugänglichkeit erfordert Wachsamkeit, weitreichende technische Fähigkeiten und leistungsfähige Computer, um diese umfangreichen Änderungen durchzuführen. Während das Archiv für digitale Medienobjekte älter wird, werden wir die Zugriffsversionen über die Jahrzehnte und Jahrhunderte sehr oft umwandeln müssen.

## Datenschutz

Wir sind als Bibliothek um den Schutz der Privatsphäre unserer Leser/innen sehr bedacht. Die Seite archive.org kann über eine verschlüsselte Verbindung aufgerufen werden und wir speichern keine IP-Adressen der Nutzer in unseren Logdateien. Wir sind der Meinung, dass die Privatsphäre der Leser wesentlicher Bestandteil der Gedanken- und Redefreiheit ist.

Das Internet Archive arbeitet aktiv mit Organisationen wie der Electronic Frontier Foundation (EFF) zusammen, um Gesetze zu unterstützen, die Nutzer schützen, und um Gesetze abzulehnen, die die Privatsphäre von Nutzern oder unsere Fähigkeit, eine öffentliche Bibliothek im Internet zu verwalten, gefährden. Im Laufe der Jahre sind wir Themen angegangen wie verwaiste Werke<sup>39</sup>, Privatsphäre<sup>40</sup> und Gesetze, die den „Safe-Harbor-Status“<sup>41</sup> für Seiten gefährden, die Nutzer-Uploads akzeptieren. Wir haben 2008 außerdem mit der Hilfe der EFF und der American Civil Liberties Union (ACLU) erfolgreich gegen einen National Security Letter gekämpft,<sup>42</sup> der von uns verlangt hätte, Informationen über Nutzer weiterzugeben.

39 Vgl. <http://blog.archive.org/2006/11/21/orphan-works-trial-nov-13th-san-francisco/> (Letzter Aufruf: 20.05.2016).

40 Vgl. <http://blog.archive.org/2014/03/11/archive-and-ala-brief-filed-in-warrantless-cell-phone-search-case/> (Letzter Aufruf: 20.05.2016).

41 Vgl. <http://blog.archive.org/2016/03/22/save-our-safe-harbor-submission-to-copyright-office-on-the-dmca-safe-harbor-for-user-contributions/> (Letzter Aufruf: 20.05.2016).

42 Vgl. <http://blog.archive.org/2008/05/16/fbi-gag-order-against-the-internet-archive-is-rescinded/> (Letzter Aufruf: 20.05.2016).



Während wir es vermeiden, persönliche Informationen über unsere Nutzer zu sammeln, wissen wir dennoch etwas über sie, da sie uns schreiben, uns Fragen stellen und mit uns über soziale Medien kommunizieren. Auskunftsbibliothekare an Universitäten nutzen [archive.org](http://archive.org), um ihren Studenten Materialien aus primären Quellen zur Verfügung zu stellen. User-Experience-Designer nutzen die Wayback Machine, um sich verändernde Trends der Gestaltung von Benutzeroberflächen im Internet zu untersuchen. Wir verfügen über eine große Community, die sich Konzerte aus unserem Archiv anhört und in unseren Foren über die besten Versionen von Songs ihrer Lieblingsbands diskutiert. Familien haben uns geschrieben, um sich für die Sicherung der Webseite eines verstorbenen Verwandten zu bedanken.

Menschen gehen aus verschiedenen Gründen in die Bibliothek – sei es aus ausbildungsrelevanten, persönlichen, geschäftlichen oder unterhaltungsbezogenen Gründen –, und die Menschen, die [archive.org](http://archive.org) nutzen, sind genauso vielfältig. Aber die Privatsphäre ist ein wesentlicher Aspekt jeder Bibliothek, und es ist unser Bestreben, dieses Ideal in der digitalen Welt beizubehalten.

## Zugänglichkeit

Die Gesellschaft entwickelt sich durch Informationen, die sie erhält, weiter; alles, was wir lernen, erfinden oder erschaffen, baut auf der Erfahrung anderer Menschen auf. Das Internet ermöglicht weltweit jedem Einzelnen Zugang zu denselben Informationen, ungeachtet seines Einkommens, Bildungsniveaus oder Wohnorts. Jeder sollte mit einem Mobiltelefon in der Lage sein, eine Bibliothek von Weltrang zu besuchen.

In diesem digitalen Zeitalter, in dem erwartet wird, dass alles online verfügbar ist, müssen wir sicherstellen, dass die besten Ressourcen zur Disposition stehen. Wir haben Jahrhunderte übergreifende wertvolle Informationen in Bibliotheken, Archiven und persönlichen Sammlungen gelagert und wir müssen sicherstellen, dass all diese Informationen online zugänglich sind. Viele Information Professionals haben ihr Leben lang auf beeindruckende Weise Wissen angesammelt und haben nun die Möglichkeit, die Früchte ihrer Arbeit mit der ganzen Welt zu teilen.

Um einen Zugang zu ermöglichen, ist es erforderlich, dass wir analoge Medien digitalisieren, Kopien von bereits digitalisierten Materialien erhalten, Methoden zur Informationsfindung entwickeln oder verbessern sowie die Medien der Öffentlichkeit zur Verfügung stellen.

Jede Bibliothek oder jedes Archiv auf der ganzen Welt wird letztendlich Fragen zum Thema Zugänglichkeit beantworten müssen. Wie können wir unsere Arbeit in einer digitalen Welt weiterführen? Wie können wir den Nutzern die Medien in einer Weise zur Verfügung stellen, dass ein Ökosystem aus Urhebern, Herausgebern, Bibliotheken und Nutzern respektiert wird? Wir haben oben einige der Methoden behandelt, die das Internet Archive nutzt, um seinen Kunden Medien online zur Verfügung zu stellen; manche Medien in unserem Archiv können kostenlos heruntergeladen werden, andere können nur gestreamt werden, einige darf man in begrenzter Menge ausleihen, bei wieder anderen können nur kurze Ausschnitte genutzt werden, und schließlich gibt es Medien, die nur sehbehinderten Menschen zum Ausdrucken zur Verfügung stehen. Über virtuelle Maschinen in unseren Datenzentren gewähren wir Forschern den kompletten Zugriff auf einige Medien, um die Untersuchung und Verarbeitung der Daten zu ermöglichen.

Eine Bibliothek erfüllt ihre Aufgabe, indem sie Medien sammelt und diese der Öffentlichkeit zur Verfügung stellt. Eine digitale Bibliothek muss in der Lage sein, dasselbe zu tun. PK

Lizenz des Textes: CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0/>)

---

## Zur Autorin

**Alexis Rossi** ist Direktorin der Abteilung Media und Access des Internet Archive. Sie arbeitet dort seit 2001, als mit der Wayback Machine der erste Dienst dieser Institution in Betrieb genommen wurde. Inzwischen verwaltet sie die Medien der gesamten Institution und stellt den Zugang sicher; dies umfasst Audiodokumente, Filme, Bücher, Software, Bilder, großflächiges Web Crawling, die Wayback Machine sowie die archive.org-Webseite. Zuvor war Alexis Rossi beim Internet Archive für die includeOpen Library und Open Content Alliance tätig.

Vor ihrer Tätigkeit für das Internet Archive verbrachte sie mehrere Jahre als Redakteurin bei ClariNet (dem ersten Online-News-Aggregator), arbeitete als Editorial Director bei Alexa Internet und als Produktmanagerin bei Mixercast. Alexis Rossi hat einen Master of Library and Information Science und hat sich auf Web-Technologien und Schnittstellen spezialisiert. Weiter liebt sie das Herstellen von Schmuck, Tanz sowie das Backen von Plätzchen.