

# Die Dark Ages des World Wide Web verhindern - Anforderungen an gelingende Webarchivierung

Jens Crueger, Digital-Historiker.de

# Geschichtswissenschaft des Digitalen

## Status quo

- Kaum Forschung mit genuin digitalen („digital born“) Quellen
- Es fehlen Methodiken und Methodologie (Quellenkritik usw.) dafür
- Institutionalisierung der digitalen Geschichtswissenschaft am Anfang
- Geschichtsschreibung der „letzten fünf Minuten“ gilt in der Geschichtswissenschaft seit jeher als heikel (Vergessensprozess nach Kosellek)

# Was uns droht...

## Dark Ages des Internet

- ☹️ Frühes WWW kaum noch erhalten
- ☹️ Durchschnittliche Bestandsdauer einer Webseite ~100 Tage
- ☹️ Dynamische Webinhalte (Nachrichtenseiten u.ä.) verändern sich alle ~7min
- ☹️ Leichte Löschbarkeit und Veränderbarkeit macht Webseiten manipulationsanfällig



Francesco Petrarca,  
1304-1374  
(Bild gemeinfrei)

# Gelingende Webarchivierung

Überlieferungschance /  
Überlieferungszufall

Systematische Archivierung /  
Sammlungsstrategie

# Träger



# Gelingende Webarchivierung

Träger

Inhalte

Formate

# Träger

„Google's mission is to organize the world's information and make it universally accessible and useful.“

# Träger

- ☺ 2001 kauft Google die Deja archives. Daraus wird Google Groups, die größte Sammlung archivierter Usenet-Nachrichten, zurückreichend bis 1981
- ☺ 2004 startet Google Books, Ziel: Jedes bekannte Buch zu retrodigitalisieren. Partnerschaften mit Bibliotheken und Entwicklung eines eigenen Buchscanners (1.000 Seiten pro Stunde).
- ☺ 2006 startet Google News Archive, Zeitungsartikel 200 Jahre zurückreichend. Seit 2008 verstärkte Retrodigitalisierung von Zeitungen.

# Träger

☹ Nach mehreren Redesigns ist Google Groups unbrauchbar zu Forschungszwecken. Bspw. Keine Suche nach Datum mehr möglich.

☹ Google News Archives gibt es seit 2011 nicht mehr.

☹ Google Books droht einzuschlafen; 2012 nur noch ca. 50% der Digitalisierungsrate von 2006; letzter Blögeintrag 2012.

☹ Google Search verlor 2011 seine Timeline Ansicht, die eine Suche nach Datum ermöglichten; Änderungen in den Algorithmen bevorzugen neuere vor älteren Webinhalten.

# Träger

- ☹ Dauerhaftigkeit privater Archive? (Negativbeispiel Google; Kopie des Internet Archive auf kanadischem Server)
- ☹ Manipulationsgefahr auch bei staatlichen Seiten (Data Rescue Events; DataRefugee.org; Schwierigkeiten bei Datenbanken)
- ☹ Kooperationschwierigkeiten (Internet Archive & LoC)

# Träger



- Eventharvesting; selektives Webharvesting; Domain.de-Crawl
- Keine gezielte Auswahl von Quellen, keine Selektionskriterien
- Private Nachrichten werden nicht archiviert
- Archivierung als Snapshot mit Metadaten
- „Pragmatische Lernschritte“

# Inhalte

Im Dezember 2014 wurde durchschnittlich 300 Stunden Videomaterial pro Minute bei der Videoplattform YouTube hochgeladen.

(Quelle: statista.com)

„Was der einen Epoche Abfall ist, ist der anderen kostbare Information.“

(A. Assmann, Erinnerungsräume, 245)

# Inhalte

- Archivierung von interpersoneller Kommunikation / Interaktion (Mails, Chats, Social Media)
- Archivierung von Egodokumenten (jenseits von Nachlässen)
- Webvideos
- Onlineshopping
- Malware
- Politisch bedeutsame Daten
- ...

# Inhalte

Randomisiertes Verfahren

Großflächige Sicherung in Reservoir

Partizipativer Ansatz

# Format



Kommunikation und Fachinformation  
für die Geschichtswissenschaften

Abonnement

Beitrag einreichen



Suchen

Home

Chancen

Forschungsberichte

Forum

Miszellen

Rezensionen

Tagungsberichte

Termine

Zeitschriften

## Snapshot

**Digital Native (Hi)Stories – Neue Herausforderungen  
audiovisuell-partizipativer Medien für die historische Bildung**

/ Termine

Drucken 

Ort	Heidelberg
Veranstaltungsort	Ruprecht-Karls-Universität Heidelberg - Historisches Seminar
Veranstalter	Studierende und Young Professionals der AG Angewandte Geschichte - Public History im Verband der Historiker und Historikerinnen Deutschlands (VHD) in Kooperation mit der Ruprecht-Karls-Universität Heidelberg und der Pädagogischen Hochschule Heidelberg
Datum	10.06.2016 - 11.06.2016
Bewerbungsschluss	05.06.2016
Url	<a href="https://www.facebook.com/agphstud/">https://www.facebook.com/agphstud/</a>

Von

**Studierende und Young Professionals der AG Angewandte Geschichte - Public History im Verband der Historiker und Historikerinnen Deutschlands (VHD)**

Kann YouTube nicht nur Unterhaltung, sondern auch Bildung? Ermöglichen solche Plattformen neue Zugangswege und die aktive Teilhabe an Geschichte? Wo liegen die

**Redaktion**

**Veröffentlicht am**

23.05.2016

**Beiträger**

Nils Steffen

**Zitation** ↓

**Klassifikation**

**Epoche**

Ohne enochalen Schwerpunkt

# Format

```

1 <!DOCTYPE html>
2 <html lang="de" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://og
3   <head>
4     <meta name="charset" charset="utf-8" />
5     <meta name="Generator" content="Drupal 8 (http://drupal.org)" />
6     <meta name="MobileOptimized" content="width" />
7     <meta name="HandheldFriendly" content="true" />
8     <meta name="viewport" content="width=device-width" />
9     <meta name="DC.Title" content="Digital Native (Hi)Stories - Neue Herausforderungen audiovisuell-partizipativer Medien für die historische Bildung" />
10    <meta name="DC.Description" content="Termin: Digital Native (Hi)Stories - Neue Herausforderungen audiovisuell-partizipativer Medien für die historische Bildung" />
11    <meta name="DC.Identifizier" content="http://www.hsozkult.de/event/id/termine-31156" />
12    <meta name="DC.Publisher" content="H-Soz-Kult" />
13    <meta name="DC.Created" content="23.05.2016" /><link href="https://fonts.googleapis.com/css?family=Cardo&subset=latin" rel="stylesheet" type="text/css" />
14    <link rel="shortcut icon" href="http://www.hsozkult.de/themes/hsk/favicon.ico" type="image/vnd.microsoft.icon" />
15    <title>Digital Native (Hi)Stories - Neue Herausforderungen audiovisuell-partizipativer Medien für die historische Bildung | H-Soz-Kult</title>
16    <link rel="stylesheet" href="http://www.hsozkult.de/core/assets/vendor/modernizr/css/modernizr.css?o81c1r" media="all" />
17    <link rel="stylesheet" href="http://www.hsozkult.de/core/modules/system/css/system.module.css?o81c1r" media="all" />
18    <link rel="stylesheet" href="http://www.hsozkult.de/modules/hfn/css/hfn.css?o81c1r" media="all" />
19    <link rel="stylesheet" href="http://www.hsozkult.de/modules/hfn/css/styles.css?o81c1r" media="all" />
20    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Source+Sans+Pro:400,400italic,600,600italic,700&subset=latin,latin-ext&o81c1r" media="all" />
21    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/mainmenu.css?o81c1r" media="all" />
22    <link rel="stylesheet" href="http://www.hsozkult.de/core/modules/user/css/user.module.css?o81c1r" media="all" />
23    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/hskfonts.css?o81c1r" media="all" />
24    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/print.css?o81c1r" media="print" />
25    <link rel="stylesheet" href="http://www.hsozkult.de/modules/hfn/css/print-preview.css?o81c1r" media="all" />
26    <link rel="stylesheet" href="http://www.hsozkult.de/core/modules/system/css/system.theme.css?o81c1r" media="all" />
27    <link rel="stylesheet" href="http://www.hsozkult.de/core/modules/filter/css/filter.caption.css?o81c1r" media="all" />
28    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/layout.css?o81c1r" media="all" />
29    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/style.css?o81c1r" media="all" />
30    <link rel="stylesheet" href="http://www.hsozkult.de/themes/hsk/css/colors.css?o81c1r" media="all" />
31
32
33 <!--[if lte IE 8]>
34 <script src="http://www.hsozkult.de/core/assets/vendor/html5shiv/html5.js?v=3.6.2"></script>
35 <![endif]-->
36 <script src="http://www.hsozkult.de/core/assets/vendor/modernizr/modernizr.min.js?v=2.6.2"></script>
37 <script src="http://www.hsozkult.de/core/assets/vendor/domready/ready.min.js?v=1.0.4"></script>
38 <script src="http://www.hsozkult.de/core/assets/vendor/jquery/jquery.js?v=2.1.0"></script>
39 <script src="http://www.hsozkult.de/core/assets/vendor/underscore/underscore.js?v=1.5.2"></script>
40 <script src="http://www.hsozkult.de/modules/hfn/js/print.js?v=1"></script>
41 <script src="http://www.hsozkult.de/modules/hfn/js/jqueryv.print-preview.js?v=1"></script>

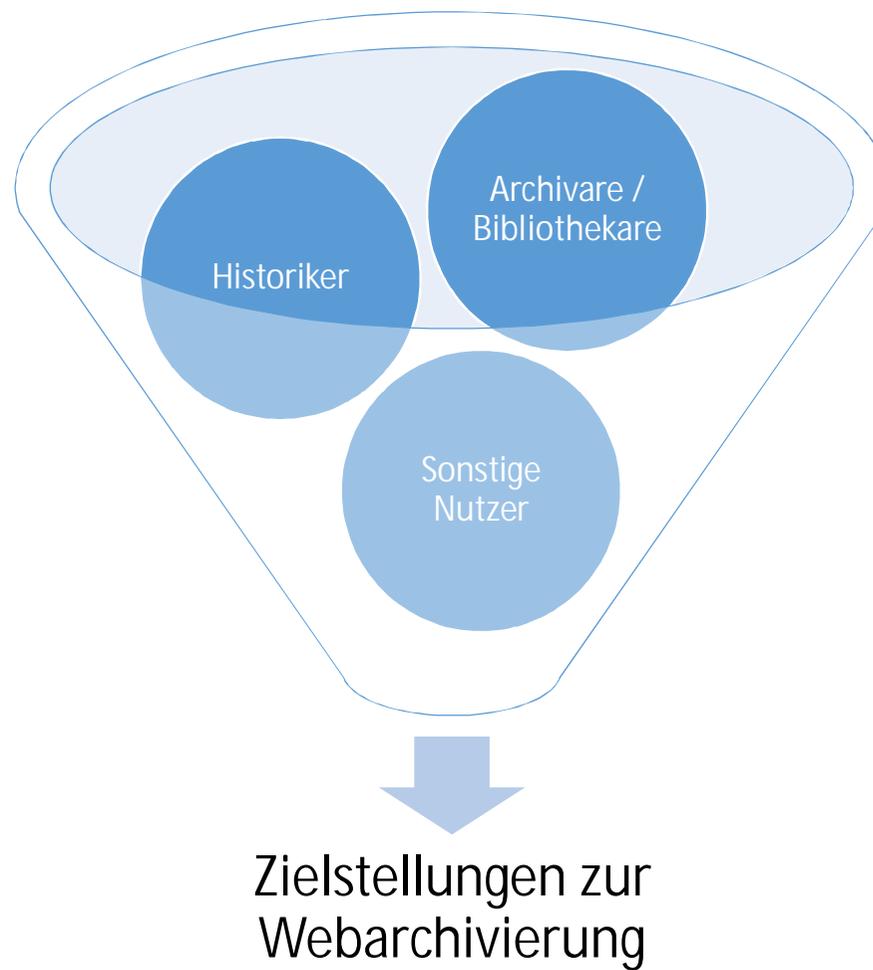
```

Quellcode

Was fehlt...

Problembewusstsein:  
Das kulturelle Gedächtnis des  
digitalen Raumes ist durch  
Quellenmangel bedroht

# Was fehlt...



**Danke für Ihre Aufmerksamkeit!**

Jens Crueger

Digital-Historiker.de