



# Gegen das digitale Vergessen

Internet Web-Inhalte werden laufend durch Neues überschrieben und verschwinden so im digitalen Nirvana. Dagegen wehren sich Archivarinnen und Aktivisten. Sie wollen Websites als Zeugnisse unserer Kultur und Geschichte erhalten.

TEXT **STEFAN MICHEL** ILLUSTRATIONEN **CHRISTINA BAERISWYL**

In den Wochen vor Donald Trumps Amtseinführung im Januar 2017 brach in Teilen der Netzgemeinde Alarmstimmung aus. Es galt, die online verfügbaren Daten der amerikanischen Umweltbehörde EPA vor Trumps Regierung zu retten. Vorausgegangen war das Gerücht, Trumps Klimawandelskeptikertruppe könnte die umfangreichen statistischen Datensätze zur globalen Erwärmung unzugänglich machen. Die Aktivisten organisierten sich an mehreren amerikanischen Universitäten und kopierten alles, was sie abrufen konnten. Genauer: Sie nutzten Computerprogramme, die Online-Inhalte kopieren, auf Server speichern und zugänglich halten. Einen grossen Teil der Seiten schickten die Aktivisten an das «Internet Archive», die Instanz schlechthin, wenn es um das Archivieren von Websites geht.

«Das Internet vergisst nicht», heisst es oft mahnend. Gemeint ist, dass Bilder, Kommentare oder Einkäufe und Suchanfragen wieder auf die Bildschirme zurückgeholt werden können, auch wenn sie eigentlich gelöscht wurden. Dies, weil gerade Aufsehenerregendes von vielen kopiert und weiterverbreitet wird. Sobald Kopien im Umlauf sind, nützt es nichts mehr, das Original zum Verschwinden zu bringen. Einen gigantischen, öffentlich zugänglichen Kurzzeitspeicher betreibt Google mit Google Cache. Dort bleiben auch ältere Versionen von Websites – und damit vielleicht der kompromittierende Social-Media-Kommentar – zwischen ein paar Tagen bis zu mehreren Wochen aufrufbar.

Menschen vergessen. Würde das Gehirn nicht laufend weniger Wichtiges aussortieren, würden wir den Überblick

verlieren. Im Internet spielt es im Prinzip keine Rolle, wie viele Inhalte gespeichert sind. Dank Suchmaschinen finden wir, was wir suchen – oder was die Suchmaschinenprogrammierer als für uns relevant erachten (aber das ist eine andere Geschichte). Inhalte können sehr wohl gespeichert und über die richtige Adresse zugänglich sein. Aber wenn diese Adresse niemandem bekannt ist, wird der Inhalt nicht gefunden. Und wenn nur schon ein Dokument von einem Ordner auf dem Server in einen anderen verschoben wird, ohne dass der Link entsprechend angepasst wird, ist ein Text, ein Bild oder eine ganze Website nicht mehr zugänglich. Der tote Link, die gekappte Verbindung ist der erste Schritt, um digital in Vergessenheit zu geraten.

Das Internet lebt in der Gegenwart. Websites verändern sich, aufgerufen wird – von Ausnahmen abgesehen – immer nur die aktuelle Version. Wie sah die Website des Tages-Anzeigers vor zehn Jahren aus? Wie vor 20 Jahren? Ziemlich sicher schlummern die alten Layouts noch auf einem Server. Hervorholen kann sie von da nur noch die Person, welche über die nötigen Server-Zugänge verfügt. Das Neue verdrängt das Alte – und nirgends läuft dieser permanente Häutungsprozess schneller als im World Wide Web.

## «Geistiges Kulturgut der Schweiz»

Das Erstaunliche dabei: Gedrucktes wird seit Jahrhunderten archiviert. Auch bei Bildern und Filmen ist man sich einig, dass die Menschen sie weiterhin sehen können sollen. Es verbindet sie mit der Vergangenheit, mit ihrer Herkunft. Texte, Fotos, Filme füttern unser kollektives Gedächtnis. Zu

diesem zählen auch Websites. Doch erst langsam setzt sich die Überzeugung durch, dass man sie ebenfalls erhalten sollte, bevor sie durch Neues überschrieben oder vom Netz genommen werden.

In der Schweiz beginnt die Zeitrechnung des Internets 2007. Aus jenem Jahr stammen die ältesten Websites, welche die Schweizerische Nationalbibliothek im «Webarchiv» abgelegt hat. Ziel dieses Archivs ist es, «eine Sammlung von landeskundlich relevanten Websites der Schweiz aufzubauen und somit das geistige Kulturgut der Schweiz langfristig zu erhalten und zur Verfügung zu stellen». Vollständigkeit strebe man nicht an, heisst es auf der Website der Nationalbibliothek. Zu begrenzt sind die Ressourcen. Die erhaltenswerten Inhalte zu finden, zu speichern und korrekt abzulegen ist Aufgabe der Kantonsbibliotheken. Ihre Mitarbeiter speisen die Inhalte in das Webarchiv des Bundes ein. Wer eine der zehn Jahre alten Websites aufrufen will, muss eine der teilnehmenden Bibliotheken besuchen. Aus Gründen des Urheberrechts dürfe man die Inhalte nur begrenzt zur Verfügung stellen, erklärt man dort auf Anfrage.

Da ist das in San Francisco betriebene «Internet Archive» um einiges offener und ambitionierter. 1996 begann der Informatiker Brewster Kahle, Schnappschüsse von Websites zu archivieren. In der «Wayback Machine» können alte Versionen und längst abgeschaltete Webpages auf den Bildschirm geholt werden. Hier findet sich beispielsweise die Online-Ausgabe des Blick vom 30. November 1996. 22 Jahre später

verzeichnet das Internet Archive nach gener Angabe 322 Milliarden Internetseiten. Die Kopierarbeit führen Computerprogramme aus. Hunderte Freiwillige bereinigen die Ablage oder suchen selber nach Inhalten, die noch nicht für die Nachwelt festgehalten wurden. Betreiber von Websites können ihre Daten auch gleich von sich aus an das Internet Archive übermitteln. Wer nun denkt, das alte Zeug interessiere doch niemanden, liegt falsch. Die Website des Internet Archive gehört zu den 300 meistbesuchten Adressen des Internets überhaupt.

### Seitenhieb gegen Google

«Wir wollen alles, das ganze verdammte Ding», wird Kahle in der Süddeutschen Zeitung zitiert. Das ist auch ein Seitenhieb gegen Google und dessen übermächtigen Suchalgorithmus. Dieser zeigt nur die Resultate an, die als relevant eingestuft werden. Dies aufgrund hunderter Kriterien, die ständig ändern. Dagegen liefert das Internet Archive beziehungsweise dessen Suchmaschine – die Wayback Machine – möglichst ungefiltert die Treffer, die den Suchbegriff enthalten. «Das ganze verdammte Ding» besitzt das gigantische Archiv dennoch nicht. Denn von den meisten Seiten wird alle zwei Monate ein Schnappschuss erstellt und abgelegt. Ein Inhalt, der nur dazwischen online war, erscheint nicht.

Doch auch mit diesem grobmaschigen Suchraster gelingen mitunter spektakuläre Treffer: beispielsweise der Post eines pro-russischen Separatistenführers in der Ukraine, in dem dieser den Abschuss eines ukrainischen Transportfluges mit Foto und Videoclip mitteilt. Keine halbe Stunde später war der Post gelöscht. In der Zwischenzeit war bekannt geworden, dass an dem Ort ein Passagierflugzeug der Malaysia Airlines mit 298 Personen an Bord abgeschossen worden war. Genau im rechten Moment hatte ein Kopierprogramm im Auftrag des Internet Archive die Website mit dem Beitrag kopiert.

Oder die US-amerikanische TV-Moderatorin Joy Reid, die bestritt, jemals homophobe Äusse-

rungen gemacht zu haben. Die Wayback Machine brachte längst gelöschte Beiträge in Joy Reids persönlichem Blog ans Licht, die zeigten, dass sie gelogen hatte. Sie zweifelte darauf die Echtheit des Archivmaterials an. Das könne ja nachträglich verändert worden sein. Später liessen Reid und ein von ihr engagierter Experte diesen Verdacht fallen.

### Lücken im Archiv

Ein Schnappschuss alle zwei Monate ist natürlich weit entfernt von einer lückenlosen Dokumentation. Noch löchriger sind die Archivbestände, wenn sie von Websites stammen, deren Inhalte erst durch Suchabfragen des Besuchers erscheinen, also zum Beispiel Online-Shops oder die Klima-Datenbank der amerikanischen Umweltbehörde. Hartwig Thomas ist ein in der Schweiz tätiger Software-Entwickler, der sich zur Internetthematik gerne öffentlich zu Wort meldet und über lange Erfahrung im Bereich der digitalen Archivierung verfügt. Er meint: «Sobald eine Website nicht einfach statisch Inhalte präsentiert, sondern auf die Besucher reagiert, haben es die Crawler schwer.» Crawler heissen die Programme, die selbständig den Inhalt von Websites untersuchen und, wenn sie Archiv-Crawler sind, kopieren und in einem Verzeichnis ablegen. Als Beispiel für Websites, die von Kopierprogrammen nur unvollständig erfasst werden, nennt Thomas die Seiten der Bundesverwaltung oder der Stadt Zürich. «Das Computerprogramm müsste Suchabfragen durchführen, damit es an den Inhalt herankäme, den es archivieren soll.» Auch die Klima-Datenbank der amerikanischen Umweltbehörde gehört in diese Kategorie.

Für diesen Fall gibt es Programme, die nicht nur kopieren, was auf einer Website angezeigt wird, sondern auch die Suchabfragen der Besucher festhalten. Das Speichern der Suchabfragen verletzt jedoch die Privatsphäre der Internetnutzer. Sinnvoller wäre es deshalb laut Thomas, Inhalte in dem Moment zu archivieren, in dem sie online gestellt werden.

«Die grössten Archivbestände hat wohl sowieso die NSA», vermutet Hernani Marques von der Netzaktivistengruppe Chaos Computer Club Schweiz. Er ist naturge-

# Texte, Fotos, Filme füttern unser kollektives Gedächtnis. Zu diesem zählen auch Websites.



mäss gegen das Speichern von Nutzerdaten. Gegen das Archivieren und zur Verfügung stellen alter Websites hat Marques hingegen nichts. «Es muss sich aber um Inhalte handeln, die für die Publikation bestimmt waren. Zudem sollten die archivierten Inhalte öffentlich zugänglich sein.»

### Die Eingreiftruppe

Manchmal muss es schnell gehen. Dann, wenn die Schliessung einer Website droht oder ihr Inhalt aus anderen Gründen gefährdet ist. Das ist besonders unerfreulich, wenn es sich um eine Plattform handelt, auf der Nutzer ihre Bilder, Texte oder andere persönliche Werke präsentieren. In solchen

Fällen ist weniger das kollektive Gedächtnis in Gefahr als vielmehr die persönliche Erinnerung. Solches zu verhindern hat sich das US-amerikanische «Archive Team» zur Aufgabe gemacht. Dessen Freiwillige verstehen sich als rebellische Eingreiftruppe und retten Daten, ob die Betreiber einer Website dies wünschen oder nicht. So hatte Nintendo vor einigen Jahren ein eigenes soziales Netzwerk namens MiiVerse geschaffen. Dort luden Spieler Bildschirmfotos und selbstgemachte Bilder hoch und tauschten sich aus. Im August 2017 kündete der japanische Spielehersteller an, dass MiiVerse im November geschlossen würde. Das Archive Team und seine Software kopierten

17 Terabytes Daten und setzten eine Archivseite auf, wo die Beiträge weiterhin angeschaut werden können.

Ist das wichtig? Für Einzelne bestimmt. Im Fall der statistischen Daten der amerikanischen Umweltbehörde definitiv. Sie wurden letztlich nicht von der EPA-Website entfernt. Sie wurden nur verschoben: dahin, wo man sie nicht mehr so leicht findet. Und für alle Fälle bleiben die Daten ausserhalb der Regierungsserver gespeichert und zugänglich. Weniger bekannte Seiten brauchen da schon mehr Glück, dass sie von einem Crawler besucht werden oder von einem Menschen vor der digitalen Apokalypse bewahrt werden.