



---

## Webarchive datenschutzgerecht füttern

### Aktuelle Kurz-Information 62

---

**Stichwörter:** Archivierung von Webseiten – Internetauftritt und Webarchive – Wayback Machine – Webarchive – Webseitenarchivierung | **Stand:** 1. April 2025

#### Was sind die Kernaussagen dieser Aktuellen Kurz-Information?

- ▶ Webarchive dokumentieren dauerhaft Versionsstände von Webseiten. Werden Webseiten archiviert, die personenbezogene Daten enthalten, ist das auch an datenschutzrechtlichen Vorgaben zu messen.
- ▶ Bayerische öffentliche Stellen sollten die Archivierung eigener Webangebote aktiv gestalten.

Webarchive haben es sich zur Aufgabe gemacht, Inhalte des Internets zu bewahren. Eines der bekanntesten Webarchive ist die sogenannte Wayback Machine. Sie ist Teil des „Internet Archive“, eines gemeinnützigen Projekts aus den USA, das sich der langfristigen Erhaltung digitaler Daten widmet. 1

Die Wayback Machine erstellt in unregelmäßigen Abständen vollständige Kopien von Internetauftritten, indem alle zugehörigen Webseiten in das Webarchiv übernommen werden. Diese „Zeitschnitte“ hält sie dann öffentlich und kostenlos zum Abruf bereit. Nutzende können etwa die Adresse (URL) einer Webseite eingeben und erhalten Informationen darüber, zu welchem Zeitpunkt diese „eingefroren“ wurde. Durch einen Klick auf das Datum kann die Webseite in ihrer jeweiligen historischen Gestalt aufgerufen werden. Dabei sind zwar nicht alle Funktionalitäten, insbesondere Links auf Unterseiten, „eingefangen“; jedoch können zahlreiche an sich längst überholte Inhalte eingesehen werden – auch zum Beispiel personenbezogene Daten von ehemaligen Beschäftigten bayerischer öffentlicher Stellen. Eine Suchfunktion über alle archivierten Webseiten, wie sie Suchmaschinen zur Verfügung stellen, hat zumindest die Wayback Machine nicht zu bieten. Allerdings wurde das Projekt bereits 1996 gegründet und hat seitdem eine beeindruckende Anzahl an Webseiten archiviert. 2

Das vorliegende Papier skizziert, was das Datenschutzrecht an Webarchiven interessant findet (1.), wirft einen Blick auf die aktuelle Entwicklung (2.) und gibt Handlungsimpulse (3.). 3

## 1. Warum interessiert sich der Datenschutz für Webarchive?

Webarchive können ein Problem für den Schutz personenbezogener Daten darstellen. Im Ausgangspunkt sind bayerische öffentliche Stellen für das Wirken unabhängiger Webarchive zwar grundsätzlich nicht datenschutzrechtlich verantwortlich.<sup>1</sup> Gleichwohl gibt es insbesondere zwei Szenarien, in denen die geschilderte Archivierung von Webseiten datenschutzrechtliche Folgefragen für bayerische öffentliche Stellen mit sich bringen kann: 4

---

### Szenario 1: Unrechtmäßig veröffentlichte personenbezogene Daten

---

So kann es vorkommen, dass versehentlich personenbezogene Daten auf einer Webseite veröffentlicht werden, die nicht zur öffentlichen Einsichtnahme bestimmt sind. Ein Beispiel 5

hierfür sind Sitzungsniederschriften aus einer nichtöffentlichen Gemeinderatssitzung, die unbeabsichtigt im Internetauftritt der Gemeinde zugänglich gemacht sind.

- 6 In der Folge crawlen Suchmaschinen die Webseiten des Internetauftritts, um Links und Vorschauen zu erstellen. Der Landesbeauftragte hat dieses Problem bereits in einem Tätigkeitsbericht aufgegriffen.<sup>2</sup> Außerdem erstellen Webarchive vollständige Kopien – einschließlich der versehentlich veröffentlichten personenbezogenen Daten.
- 7 Während die Daten von der ursprünglichen Webseite relativ unkompliziert und zeitnah gelöscht werden können und sich die Suchergebnisse entsprechend – wenn auch mit Verzögerung – aktualisieren, verschwinden die Daten aus Webarchiven nicht ohne weiteres Zutun. Ohne Rechtsgrundlage veröffentlichte personenbezogene Daten können so dauerhaft in Webarchiven sichtbar und abrufbar bleiben. Dies erschwert es einer öffentlichen Stelle, die Folgen eines Datenschutzverstoßes – hier einer versehentlichen Veröffentlichung personenbezogener Daten – zu beseitigen bzw. einzudämmen.

---

### Szenario 2: Veraltete personenbezogene Daten

---

- 8 In einem weiteren Szenario werden personenbezogene Daten zunächst mit Rechtsgrundlage auf einer Webseite veröffentlicht – beispielsweise eine Liste mit wichtigen Ansprechpersonen für Bürgerinnen und Bürger bei einer Gemeinde. Das Angebot wird in der Folge aber sukzessive bei personellen Wechseln nachgeführt.
- 9 Auch hier werden Suchmaschinen und Webarchive aktiv. Während sich die Suchergebnisse von Suchmaschinen meist ausreichend schnell automatisch aktualisieren, entfernen sich die „eingefrorenen“ Kopien in den Webarchiven mit der Zeit immer weiter von der aktuellen Lage. Jedenfalls die Gemeinde dürfte beispielsweise veraltete Ansprechpersonen-Listen nicht mehr öffentlich bereithalten, weil dies zu ihrer Aufgabenerfüllung nicht (mehr) erforderlich ist.

## 2. Wie schaut die aktuelle Entwicklung aus?

- 10 Aus Datenschutzsicht ist es daher erfreulich, dass Google im Februar 2024 mitgeteilt hat, den Google-Cache einzustellen. Die Google-Cache-Funktion diene dazu, Momentaufnahmen von Webseiten zu speichern, wie sie zum Zeitpunkt der letzten Google-Indexierung aussahen. Im Gegensatz zu Webarchiven, die verschiedene Versionen einer Seite speichern, hielt der Google-Cache lediglich eine, oft relativ aktuelle Kopie bereit. Selbst dies ermöglichte aber mitunter den Zugriff auf veraltete Daten.
- 11 Mittlerweile verweist Google allerdings stattdessen auf die Wayback Machine. Im September 2024 gab das „Internet Archive“ zudem bekannt, dass der Blick in die Vergangenheit des Internets mit der Wayback Machine nun leichter zugänglich sei als bisher<sup>3</sup>. Über die Google-Suchergebnisse könne nicht nur der Link auf die aktuelle Webseite gefunden werden, sondern – mit nur wenigen Klicks – auch ein Weg zu alten Versionen genau dieser Webseite.
- 12 Während bisher die genaue URL einer Webseite bekannt sein musste, um auf alte Versionen zugreifen zu können, ist es nun möglich, über die Eingabe von Suchbegriffen in Google – ausgehend von der aktuellen Version der Webseite – zu diesen alten Versionen zu gelangen.

Google vorschlagwortet dabei die Wayback Machine selbst aber nicht. Das bedeutet, dass eine aktive Suche nach veralteten Daten nicht möglich ist. Dennoch wird der Zugang zu veralteten personenbezogenen Daten über die Wayback Machine deutlich erleichtert.

### 3. Gibt es nun etwas zu tun?

Die Veröffentlichung personenbezogener Daten auf Webseiten stellt eine Verarbeitung dieser Daten nach Art. 4 Nr. 2 Datenschutz-Grundverordnung (DSGVO) dar. Bei jeder Verarbeitung personenbezogener Daten haben Verantwortliche die Datenschutzgrundsätze nach Art. 5 Abs. 1 DSGVO zu beachten; Verantwortliche müssen dies auch nachweisen können (Rechenschaftspflicht, Art. 5 Abs. 2 DSGVO). **13**

Will eine öffentliche Stelle personenbezogene Inhalte in ihrem Internetauftritt einstellen, muss sie sich somit – wie stets – zunächst vergewissern, dass diese Verarbeitung von einer hinreichenden Rechtsgrundlage gedeckt ist (Grundsatz der Rechtmäßigkeit, Art. 5 Abs. 1 Buchst. a, Art. 6 Abs. 1 DSGVO). Besondere Bedeutung erlangen dabei die „Erforderlichkeit“ der Veröffentlichung sowie der Grundsatz der Datenminimierung (Art. 5 Abs. 1 Buchst. c DSGVO): Eine Veröffentlichung personenbezogener Daten auf einer Webseite ist ohnehin nur in dem Umfang zulässig, wie dies zur Erreichung des jeweiligen Verarbeitungszwecks notwendig und angemessen ist. Als – aus Datenschutzsicht positiver – „Nebeneffekt“ kommt im vorliegenden Zusammenhang für die öffentliche Stelle hinzu: Je weniger personenbezogene Daten auf einer Webseite veröffentlicht werden, desto weniger Sorge bereiten auch unerkannte oder gegebenenfalls unerlaubte Kopien. **14**

Verantwortliche haben ferner durch geeignete technische und organisatorische Maßnahmen nachweisbar sicherzustellen, dass eine Verarbeitung personenbezogener Daten im Einklang mit der Datenschutz-Grundverordnung erfolgt (Art. 24 Abs. 1 Satz 1 DSGVO). Insbesondere ist durch technische und organisatorische Maßnahmen zu gewährleisten, dass die Datenschutzgrundsätze – wie etwa der Grundsatz der Datenminimierung – wirksam umgesetzt werden (Art. 25 Abs. 1 DSGVO). Zu bedenken ist auch: Hat ein Verantwortlicher personenbezogene Daten öffentlich gemacht und ist er nach Art. 17 Abs. 1 DSGVO zu deren Löschung verpflichtet, muss er angemessene Maßnahmen treffen, um andere Verantwortliche, welche die betreffenden Daten verarbeiten, über das Lösungsverlangen der betroffenen Person zu informieren (Art. 17 Abs. 2 DSGVO).<sup>4</sup> **15**

Öffentliche Stellen sollten daher stets kritisch prüfen, ob und in welchem Umfang sie personenbezogene Daten im eigenen Internetauftritt veröffentlichen dürfen. Eine solche Überprüfung ist eine Daueraufgabe: Veröffentlichte Daten können nämlich veralten, und eine einstmals zulässige Veröffentlichung dieser Daten kann mangels Erforderlichkeit durch Zeitablauf unzulässig werden. Eine dergestalt datensparsame Gestaltung des Internetauftritts lohnt sich auch deshalb, weil sie hilft, Ärger zu vermeiden, der sich aus dem Wirken von Webarchiven ergeben kann. **16**

Gegebenenfalls kommen öffentliche Stellen auch zu dem Schluss, dass sie eine Archivierung durch ein Webarchiv generell unterbinden und bereits bestehende Kopien löschen lassen wollen. Soll eine Archivierung nicht gänzlich verhindert werden, ist es ratsam, den Internetauftritt auf möglicherweise problematische Unterseiten mit personenbezogenen Daten zu **17**

überprüfen. Werden solche gefunden, sollte versucht werden, den Zugriff durch das Webarchiv für diese Unterseiten zu blockieren, um eine Archivierung zu vermeiden.

- 18 Um die Löschung älterer Versionen zu veranlassen und die Erstellung zusätzlicher Kopien zu vermeiden, empfiehlt etwa die Wayback Machine in ihren häufig gestellten Fragen<sup>5</sup>, eine E-Mail an [info@archive.org](mailto:info@archive.org) zu senden. Darin sollten folgende Informationen enthalten sein:
  - die URL der betreffenden Webseite,
  - der Zeitraum, der von der Archivierung ausgeschlossen werden soll,
  - der Zeitraum, in dem die öffentliche Stelle die Kontrolle über die Webseite hatte,
  - zusätzliche Informationen, welche die Anfrage besser verstehen helfen.
- 19 Im Internet sind zudem Erfahrungsberichte zu alternativen Methoden zu finden, wie eine Archivierung durch Webarchive möglicherweise verhindert werden kann:
  - So soll ein Eintrag in der Datei „robots.txt“ mit dem Inhalt „User-agent: archive.org\_bot-Disallow: /“ eine Nichtarchivierung bewirken können.
  - Wird ein Apache-Webserver verwendet, soll über die Datei „htaccess“ ein Zugriff für Webarchive blockiert werden können. Dazu müssen entweder die IP-Adressen der Webarchive oder der „User-Agent-String“ bekannt sein. Diese Methode könne auch dazu beitragen, Webarchive daran zu hindern, bestimmte Unterseiten zu archivieren.
- 20 Der Landesbeauftragte hat diese Methoden nicht getestet; sie sind hier lediglich informativ aufgeführt und mögen auch in Betracht kommen, wenn der von der Wayback Machine empfohlene Ansatz nicht erfolgreich sein sollte.
- 21 Erfahrungsberichte sprechen dafür, dass sich Webarchive nicht immer zuverlässig blockieren lassen. Auch vor diesem Hintergrund kommt einer bewusst datensparsamen Gestaltung von Webangeboten besondere Bedeutung zu.

<sup>1</sup> Zu beachten ist allerdings, dass der Europäische Gerichtshof den Begriff des Verantwortlichen weit auslegt und in der Folge die Schwelle für eine gemeinsame Verantwortlichkeit recht niedrig ansetzt (vgl. nur Europäischer Gerichtshof, Urteil vom 5. Juni 2018, C-210/16, Rn. 26 ff.). Vor diesem Hintergrund kommt bei einer Archivierung von Webseiten insbesondere im Falle eines „Zusammenwirkens“ einer öffentlichen Stelle mit einem Webarchiv gegebenenfalls eine (gemeinsame) Verantwortlichkeit auch der öffentlichen Stelle in Betracht. Maßgebend sind insoweit die konkreten Umstände des Einzelfalls. Näher hierzu Bayerischer Landesbeauftragter für den Datenschutz, Gemeinsame Verantwortlichkeit, Orientierungshilfe, Stand 6/2024, Internet: <https://www.datenschutz-bayern.de/infothek>.

<sup>2</sup> Bayerischer Landesbeauftragter für den Datenschutz, 32. Tätigkeitsbericht 2022, Nr. 1 2.4.1.

<sup>3</sup> Siehe dazu <https://blog.archive.org/2024/09/11/new-feature-alert-access-archived-webpages-directly-through-google-search/>, zuletzt abgerufen am 10. Januar 2024.

<sup>4</sup> Ausführlich zu diesem „Recht auf Vergessenwerden“ Bayerischer Landesbeauftragter für den Datenschutz, Das Recht auf Löschung nach der Datenschutz-Grundverordnung, Orientierungshilfe, Stand 6/2022, Rn. 60 ff., Internet: <https://www.datenschutz-bayern.de/infothek>.

<sup>5</sup> Siehe dazu <https://help.archive.org/help/how-do-i-request-to-remove-something-from-archive-org/>, zuletzt abgerufen am 10. Januar 2024.