KULTURGUT WEB

Michaela Mayr

1. Kulturerbe einst und jetzt

Die historischen Wurzeln der Österreichischen Nationalbibliothek reichen weit in die Geschichte zurück. Aus dem Jahre 1368 stammt das erste nachweisbare, noch heute in der Bibliothek vorhandene Buch, das so genannte „Evangeliar des Johannes von Troppau", eine mittelalterliche Prachthandschrift. Von der Hofbibliothek des Habsburgischen Kaiserreiches bis zur Österreichischen Nationalbibliothek der Neuzeit wurden Millionen von physischen Objekten (vor allem Druckwerke, aber auch eine Vielzahl von Manuskripten, Bildern und Porträts, Karten, Papyri und vieles mehr) gesammelt, die die Bibliothek zu einer zentralen Gedächtnisinstitution des Landes anwachsen ließen.

Die Österreichische Nationalbibliothek sieht sich selbst als Schnittstelle zwischen ihrem reichen historischen Kulturerbe und einer dynamischen Wissensgesellschaft von morgen. Ihr zentraler Platz in einer sich rasch verändernden digitalen Medienwelt erfordert das permanente Hinterfragen von Zielsetzungen und Prozessen und das Anpassen an neue Medienformen und Erwartungen von BenützerInnen. Sowohl in der zukunftsweisenden Vision 2025[1], als auch in der konkreten Strategie 2012–2016[2] wird dieser Entwicklung Rechnung getragen: Digitalisierung der Bestände, Langzeitarchivierung digitaler Medien, vereinfachter Zugang zum Wissen durch zeitgemäße Technologien und digitale Services für die Forschung sind zentrale Bestrebungen für die Zukunft der Österreichischen Nationalbibliothek.

Erfolgreiche, langjährige Digitalisierungsinitiativen von Beständen bestätigen die Richtigkeit und Bedeutsamkeit dieses Weges. Seit über zehn Jahren stellt das Projekt ANNO (AustriaN Newspapers Online)[3] einen virtuellen Zeitungslesesaal mit mehr als 15,5 Millionen Seiten von historischen österreichischen Zeitungen und Zeitschriften zu Verfügung. Im Rahmen von Austrian Books Online digitalisiert die Österreichische Nationalbibliothek ihren kompletten historischen Buchbestand vom 16. bis in die zweite Hälfte des 19. Jahrhunderts (600 000 urheberrechtsfreie Werke) und macht diesen sukzessive online zugänglich und im Volltext durchsuchbar.

---

1 Österreichische Nationalbibliothek Wien, Vision 2025 – Wissen für die Welt von morgen. http://www.onb.ac.at/about/21043.htm, zuletzt abgerufen am 11.5.2016.

2 Österreichische Nationalbibliothek Wien, Strategie 2012 – 2016. http://www.onb.ac.at/about/22407.htm, zuletzt abgerufen am 11.5.2016.

3 Österreichische Nationalbibliothek Wien, ANNO (AustriaN Newspapers Online). http://anno.onb.ac.at/, zuletzt abgerufen am 11.5.2016.

Mittlerweile ist ein immer größerer Teil der produzierten Information digital und Gedächtnisinstitutionen sehen sich mit der Herausforderung konfrontiert, auch dieses digitale Wissen für die Zukunft zu sichern. Das österreichische Mediengesetz[4] regelt die Anbietungs- bzw. Abgabepflicht von „Bibliotheksstücken" und wurde im Jahr 2000 von „Druckwerken" auf „sonstige Medienwerke" (einschließlich Offline-Publikationen, d.h. Medien auf einem festen Datenträger, z.B. CD-ROMs, DVDs) erweitert. Eine neuerliche Mediengesetznovelle, die am 1. März 2009 in Kraft trat, ermächtigt die Österreichische Nationalbibliothek, auch Online-Publikationen zu sammeln und ein Archiv österreichischer Websites zu betreiben.[5] Seither erstreckt sich der Sammelauftrag der Österreichischen Nationalbibliothek auch auf das digitale Kulturerbe des Landes.

2. Webarchivierung

Im Jahr 1996 erkannte der amerikanische Informatiker und Unternehmer Brewster Kahle das Desiderat, Webseiten langfristig für die Nachwelt zu erhalten und gründete in San Francisco das Internet Archive.[6] Seither wurden von der Non-Profit-Organisation 484 Milliarden Webseiten archiviert und online zugänglich gemacht.[7]

Zahlreiche internationale Initiativen folgten (Vorreiter waren z.B. skandinavische Nationalbibliotheken sowie Australien) und zwölf Institutionen (inklusive Internet Archive) riefen schließlich 2003 das International Internet Preservation Consortium (IIPC)[8] ins Leben. Aktuell beteiligen sich neben der Österreichischen Nationalbibliothek Organisationen aus über 45 Ländern, in erster Linie Universitäten, Bibliotheken und Archive, an den Aktivitäten des Konsortiums.

Das Web@rchiv Österreich wurde im Jahr 2008 an der Österreichischen Nationalbibliothek gegründet. Nach Aufbau von Organisation und Infrastruktur begann ab Inkrafttreten der Mediengesetznovelle im März 2009 der operative Betrieb der Archivierung.

Der Zugriff auf das Webarchiv ist ebenfalls durch das Mediengesetz geregelt und kann vor Ort in der Österreichischen Nationalbibliothek und bei ausgewählten berechtigten Bibliotheken, nicht aber online, erfolgen. Die elektronische Weiterverarbeitung

[4] Österreichisches Mediengesetz: https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=BgblAuth&Dokumentnummer=BGBLA_2009_I_8, zuletzt abgerufen am 11.5.2016.

[5] Vgl. Bettina Kann, Webarchiv Österreich: Digitales Wissen sichern. In: BIBLIOTHEK Forschung und Praxis, Band 35, Heft 1. Seiten 26-32. http://dx.doi.org/10.1515/bfup.2011.004, zuletzt abgerufen am 11.5.2016.

[6] Internet Archive: https://archive.org/, zuletzt abgerufen am 11.5.2016.

[7] Internet Archive Wayback Machine: https://archive.org/web/, zuletzt abgerufen am 11.5.2016.

[8] International Internet Preservation Consortium: http://www.netpreserve.org/, zuletzt abgerufen am 111.5.2016.

der archivierten Daten ist nicht gestattet, lediglich ein Ausdruck der Webseiten kann bei Bedarf erstellt werden.

2.1. Eckdaten des Web@rchiv Österreich

Seit Beginn der Sammelaktivitäten im Jahr 2009 wurden vom Web@rchiv Österreich online Inhalte im Ausmaß von rund 84 Terabyte (Stand Mai 2016) gespeichert. Der Prozess des Sammelns, das sogenannte Harvesting oder Crawling, wird inhouse durchgeführt, die physische Speicherung der Daten erfolgt im Bundesrechenzentrum. Die archivierten Medien setzen sich aus 2,7 Milliarden Einzeldateien zusammen und stammen von ca. 1,7 Millionen Domains. Der Großteil der Domains entfällt dabei auf die .at Domäne inklusive der Subdomains .ac.at (Wissenschaft, Bildung) und gv.at (Verwaltung, Behörden), und inkludiert seit kurzem auch die neue Top Level Domain .wien.

Der operative Betrieb des Webarchivs wird von zwei MitarbeiterInnen in der Digitalen Bibliothek koordiniert. Harvesting und Zugriff auf die Webseiten erfolgt mittels Open Source Software Heritrix[9], NetarchiveSuite[10] und (Open) Wayback Machine[11], die aus internationalen Kooperationen hervorgegangen sind.

2.2. Modelle zur Webarchivierung

Durch eine Kombination unterschiedlicher Strategien und Archivierungsansätze sollen möglichst aussagekräftige Momentaufnahmen des österreichischen Webspace geschaffen und für die Nachwelt festgehalten werden:

2.2.1. Domain Harvesting

Diese breit angelegten Crawls, in der Fachsprache auch als „snapshot harvesting" oder „broad crawls" bezeichnet, decken zumeist ganze Domänen ab. Sie liefern Momentaufnahmen einer großen Anzahl von Webseiten, gehen aber üblicherweise nicht in die Tiefe. Der kuratorische Aufwand ist dabei gering, die Methode zielt nicht auf konkrete inhaltliche Seitenauswahl oder Bewertung, sondern auf Masse ab, und ist daher mit einem hohen technischen Ressourceneinsatz verbunden.

[9] Internetarchive/Heritrix3: https://github.com/internetarchive/heritrix3, zuletzt abgerufen am 11.5.2016.

[10] NetarchiveSuite: https://sbforge.org/display/NAS/NetarchiveSuite;jsessionid=1A57DA390410BEB4C763868511542E77, zuletzt abgerufen am 11.5.2016.

[11] Open Wayback Machine: https://github.com/iipc/openwayback/wiki, zuletzt abgerufen am 11.5.2016.

Die österreichische Top Level Domain.at umfasst aktuell (Stand Mai 2016) nahezu 1,3 Millionen Domains.[12] Das Web@rchiv Österreich archiviert darüber hinaus auch Inhalte anderer Top Level Domänen mit Österreich Bezug wie z.B. com, .net, .org, .info, .cc, .eu etc.. In diesen Fällen ist allerdings meist eine aufwändige manuelle Auswahl erforderlich. Gerne werden Seitennominierungen von BibliothekarInnen, BenutzerInnen oder SeitenbetreiberInnen aufgenommen. Mit der Schaffung neuer Top Level Domains wie z.B. .wien oder .tirol wächst der österreichische Webspace und somit das Web@rchiv Österreich um weitere Adressen an.

Alle zwei Jahre finden diese umfangreichen Domain Crawls durch das Web@rchiv Österreich statt, 2015 bereits zum vierten Mal.

2.2.2. Selektives und Event Harvesting

Aufgrund der langen Intervalle zwischen den Domain Crawls und dem damit verbundenen Risiko, dass in der Zwischenzeit zahlreiche Online Publikationen verloren gehen, werden für ausgewählte Seiten, die häufigen Änderungen unterliegen oder thematisch von besonderem Interesse sind, in kürzeren Abständen selektive Harvestings durchgeführt. Webseiten zu speziellen Anlässen und Großereignissen (z.B. Wahlen, Sportveranstaltungen wie Olympische Spiele etc.) sind besonders gefährdet und stehen meist nur für den Zeitraum des Ereignisses zur Verfügung.

Die Auswahl der Webseiten, die Analyse der Struktur und die Festlegung individueller Archivierungsintervalle verursachen einen hohen kuratorischen Aufwand, ermöglichen aber andererseits eine zielgerichtete Speicherung und höhere Qualität durch maßgeschneiderte Archivierungsparameter. Anfangs sind zusätzlich administrative Aufgaben zu verrichten, da das Mediengesetz in Österreich vorsieht SeitenbetreiberInnen vor Durchführung der Archivierung verpflichtend darüber zu informieren.

Selektive Harvestings können kontinuierlich oder zeitlich befristet ausgeführt werden. Das Web@rchiv Österreich betreibt z.B. eine laufende Medien- und Politikkollektion auf Basis täglicher Crawls. Zeitlich abgegrenzte thematische Crawls erstreckten sich beispielsweise auf Inhalte zum Gedenkjahr Erster Weltkrieg 2014, oder als Event Crawl zum Eurovision Song Contest 2015.

Überwiegend werden selektive oder Event Crawls vorab geplant, es kann aber durchaus erforderlich sein, spontan zu reagieren und im Bedarfsfall ungeplante Harvestings zu starten, wenn bedeutsame Ereignisse wie beispielsweise Naturkatastrophen oder Terroranschläge (z.B. 9/11 New York, Charlie Hebdo Paris) etc. eintreten.

[12] Nic.at: Domain Statistiken: https://www.nic.at/uebernic/statistiken/, zuletzt abgerufen am 11.5.2016.

2.3. Herausforderungen bei der Archivierung von Webseiten

Die Archivierung von Webseiten lässt sich in Bezug auf ihre Methodik nicht mit der Sammlung von analogen Materialien vergleichen. Born digital Medien (insbesondere Webseiten) weisen Charakteristika auf, die digitale Bibliotheken und Archive vor besondere Herausforderungen stellen:

Ungeheure elektronische Datenmengen kommen auf Gedächtnisinstitutionen zu. Laut EMC Studie von 2014[13] verdoppelt sich das digitale Universum alle zwei Jahre. Für 2020 werden 44 Zettabytes an Daten prognostiziert. Von 2013 bis 2020 wird die Menge von 4,4 Zettabytes um den Faktor 10 anwachsen. Bereits jetzt existieren im deutschsprachigen D-A-CH Raum insgesamt 19 Millionen .de, .at und .ch Domains, die teilweise von den jeweiligen Nationalbibliotheken archiviert werden. Sowohl die Anzahl der Domains, als auch die Speichermengen der jeweiligen Seiten steigen kontinuierlich an und bedingen dadurch einen erhöhten Ressourcenbedarf bei Webarchiven.

Die Kurzlebigkeit von Inhalten im World Wide Web erfordert von Webarchiven häufig Flexibilität und rasche Reaktionsfähigkeit. Gilt es doch, manche Seiten (z.B. von zeitlich begrenzten Events) binnen sehr kurzer Frist zu speichern, bevor sie für immer verloren gehen.

Was die Archivierungs- und Zugangsmethoden anbelangt, ist es für Webarchive oft schwierig mit der dynamischen technologischen Entwicklung im Bereich der Online Medien Schritt zu halten. Daraus ergeben sich zahlreiche technische Problemstellungen, die in den meist kleinen Teams und Einzelprojekten nicht gelöst werden können. Trotz internationaler Kooperationen, wie dem International Internet Preservation Consortium, gelingt es nach wie vor nur bedingt, alle gewünschten Webseiten vollständig zu archivieren bzw. realitätsgetreu wiederzugeben.

Unterschiedliche nationale Gesetzgebungen regeln mehr oder weniger genau, ob und wie die Speicherung von Webseiten und der Zugriff auf die Archive in einem Land erlaubt sind. Wie kann man dabei nationale Grenzen ziehen, was ist überhaupt das „österreichische" Internet? Es gibt verschiedene Ansätze, die dabei verfolgt werden: Zur Abgrenzung können beispielsweise nationale Top Level Domains wie .at, .de, .ch etc. herangezogen werden. Wie aber sind generische Domains (z.B. .com, .net, info etc.) zu behandeln? Der Standort des Servers oder inhaltliche Kriterien wie Sprache, Zielpublikum, Adressangaben im Impressum usw. dienen oftmals als weitere Unterscheidungsmerkmale. Durch die faktische Grenzenlosigkeit des World Wide Web bewegen sich Webarchive allerdings teilweise in einem rechtlichen Graubereich.

[13] EMC Digital Universe with Research & Analysis by IDC: The Digital Universe of Opportunities. http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm, zuletzt abgerufen am 11.5.2016.

Mit dem Einzug von Web 2.0 und sozialen Medien wie Facebook, Twitter etc. ist auch in der Webarchivierung eine neue Ära angebrochen. Der interaktive Charakter der Sozialen Netzwerke stellt Webarchive technisch und rechtlich vor neue, noch komplexere Aufgaben.

3. Fazit und Ausblick

Nach Jahren intensiver Aufbauarbeit und Weiterentwicklung der Archivierungsmethoden hat sich die Webarchivierung international als Disziplin etabliert. Obwohl viele Initiativen nach wie vor über einen geringen Bekanntheitsgrad verfügen (meist bedingt durch eingeschränkten offline Zugang), ermöglichen Leuchtturmprojekte wie das Internet Archive die Sichtbarkeit von Webarchiven und verdeutlichen deren Sinnhaftigkeit. Die aktuelle Diskussion in der Community spiegelt zusehends den Bedarf der Einbindung von Forschung und Wissenschaft wider.

Ein wichtiges Desiderat für die Zukunft wären verstärkte internationale Kooperationen und zentral nutzbare online Webarchive. Dies scheitert aktuell vor allem an unterschiedlichen nationalen Gesetzgebungen und zahlreichen Regelungen, die den online Zugriff einschränken.

Eine weitere essentielle Herausforderung für die Zukunft ist die Langzeitarchivierung von archivierten Online Medien. Es muss sichergestellt werden, dass Generationen von künftigen WissenschafterInnen wichtige Quellen zu unserer Gegenwart erforschen können.

ÜBERLEGUNGEN ZUR ÜBERNAHME UND ARCHIVIERUNG VON E-MAIL-KONTEN

Corinna Knobloch

Die Frage der Übernahme von E-Mail-Konten bzw. von einzelnen E-Mails wird in der digitalen Archivierung nun schon seit einigen Jahren nur am Rande thematisiert.[1] Und dennoch: Die Frage der Erhaltung von E-Mail-Ablagen über die Lebenszyklen von Hard- und Software hinweg gewinnt immer mehr an Bedeutung, da E-Mails im Laufe der letzten 15 Jahre einen deutlichen Funktionswandel erfahren haben. Inzwischen kann man längst nicht mehr davon ausgehen, dass alle wichtigen E-Mails ausgedruckt Eingang in die Papierakten finden.

1. Zur Funktion von E-Mails

Zunächst ist aus der archivischen Sicht zu klären, welche Funktionen E-Mails haben und welche Konsequenzen daraus für die Archivierung abzuleiten sind.

Die Spanne ist groß:

- Oftmals erfüllen die E-Mails Funktionen, die zuvor vorwiegend von Briefen, Aktenvermerken und anderen Schreiben wahrgenommen worden sind. Sie können diese Funktionen einerseits eigenständig leisten, andererseits aber auch eine Ergänzung zu Papierschriftstücken und anderen Unterlagen sein. In manchen Fällen – insbesondere wenn keine konsequente Aktenführung in der Behörde stattgefunden hat – stellen sie in ihrer elektronischen Form die einzige aussagekräftige Aufzeichnung des Verwaltungshandelns dar. Das ist dann der Fall, wenn die handelnden Personen die relevanten E-Mails nicht ausgedruckt und zu den Akten gegeben haben.
- Auf der anderen Seite kann die E-Mail ein schnelles Austauschmedium sein. Sie dient dann beispielsweise zur Absprache von Terminen und Arbeitsschritten und übernimmt damit die kurzlebigen Funktionen von Notizzetteln, Anrufen oder anderen Gesprächen.

Aufgrund der unterschiedlichen Funktionen ist bei der Bewertung von E-Mail-Konten wie bei den meisten anderen behördlichen Unterlagen davon auszugehen, dass einige Teile archivwürdig sein können, andere Teile hingegen nicht.

[1] Als Beispiel sei an dieser Stelle nur genannt: Mike Zuchet, Pilotprojekt zur Langzeitarchivierung digitaler E-Mail-Korrespondenzen des Bundesvorstands der Vereinigten Dienstleistungsgewerkschaft ver.di. In: Christian Keitel – Kai Naumann (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen". Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Serie A, 24. Stuttgart 2013. S. 165-170.

MITTEILUNGEN DES ÖSTERREICHISCHEN STAATSARCHIVS

DIGITALE ARCHIVIERUNG

Innovationen – Strategien – Netzwerke

HERAUSGEGEBEN VON DER GENERALDIREKTION

59

2016