



Das Luxemburger Webarchiv

BnL – Wëssen entdecken (12): Wozu und wie werden Internetseiten aufbewahrt?

Von Ben Els*

Auf den Fehlercode 404 stoßen wir, wenn Links nicht mehr funktionieren, eine Seite verschwunden ist oder die Internetadresse gewechselt hat. Bilder, Videos, Dokumente und Informationen, auf die wir gestern noch freien Zugriff hatten, können heute bereits verschwunden sein. Diese Enttäuschung auf Seiten der Nutzer steht in keinem Verhältnis zur Frustration auf Seiten der Produzenten: Wenn eines Tages alle Inhalte und persönlichen Beiträge von einer Plattform verschwunden sind, bleibt nichts weiteres als der Ärger über fehlende Kopien auf dem eigenen Rechner.

Zu den Kernaufgaben der Nationalbibliothek gehört die Aufbewahrung von Publikationen jeder Art. Allerdings ist das Bewusstsein über diese Abgabepflicht, dem „dépôt légal“, gerade im digitalen Bereich, und insbesondere in Bezug auf die Archivierung ganzer Webseiten wenig verbreitet.

Das Luxemburger Web wird von der BnL im nationalen Webarchiv zusammengetragen und aufbewahrt. Regelmäßig versucht sie, alle in Luxemburg veröffentlichten Seiten sowie alle als .lu registrierten Domains (Adressen die auf .lu enden) aufzuspüren und zu archivieren. Das bedeutet, dass Suchroboter, die sogenannten „Webcrawler“, alle Elemente dieser Seiten durchlaufen, abspeichern und daraus exakte Kopien anfertigen, die alle Dokumente und Medien sowie die Funktionsweise des Originals beinhalten. Im Laufe der Zeit entsteht aus mehreren Momentaufnahmen eine Chronologie der gleichen Webseite, durch die man in der Zeit

vor- und zurückspringen kann. Auf diesem Weg lässt sich die Evolution einer Webseite zurückverfolgen sowie bereits gelöschte Webseiten wiederfinden. So wurde etwa festgestellt, dass innerhalb von drei Jahren rund 20 000 registrierte .lu Domains nicht mehr verfügbar waren. Auch wenn die Gesamtzahl von registrierten Domains seit Jahren stetig steigt, veranschaulicht diese Fluktuation bei ungefähr 100 000 .lu Domains den rasanten Wandel im Internet.

Diese Wechselhaftigkeit hat als Konsequenz, dass die BnL nicht in der Lage ist, jedes Update jeder Seite zu jedem Zeitpunkt einzufangen. Das Webarchiv muss also die zur Verfügung stehenden Ressourcen effektiv einsetzen, um so viele Inhalte als möglich aufbewahren zu können. Aus diesem Grund werden verschiedene Herangehensweisen in der Archivierung angewandt: 1. regelmäßige „Domain Crawls“ als breitflächige Aufnahme des gesamten erkundeten Luxemburger Webs; 2. Event-Kollektionen, in denen über einen bestimmten Zeitraum ein Maximum an Informationen zu Ereignissen nationaler Bedeutung gesammelt wird; 3. Thematische Kollektionen, deren Listen relevanter Webseiten sich mit der Zeit weiterentwickeln und bei deren Zusammenstellung das Webarchiv verstärkt auf externe Unterstützung setzen möchte.

Über das Informationsportal webarchive.lu finden Besitzer und Nutzer des Luxemburger Webs alle nötigen Informationen zu Aufgaben und Funktionsweise des Webarchivs. Hier präsentiert die Nationalbibliothek die Resultate

bisheriger Event-Kollektionen, wie zum Beispiel der verschiedenen politischen Wahlkampagnen zwischen 2017 und 2019 sowie der Covid-19-Sammlung, die seit März 2020 fortwährend Informationen von hunderten Webseiten mit Bezug zur Pandemie in Luxemburg einsammelt.

Ein wichtiger Aspekt der Arbeit des Webarchivs besteht darin, zur Teilnahme an der Bewahrung des Luxemburger Webs mitzuwirken: mit Hilfe des Kontaktformulars oder einer einfachen E-Mail können Vorschläge für neue Einträge und Ergänzungen unserer Archiv-Listen eingesandt werden. Gerade im Bereich der Webseiten, die nicht als .lu-Domain registriert sind (also alle luxemburgischen Seiten die auf .com, .de, .fr etc. enden), ist die Nationalbibliothek auf Hilfe und Hinweise angewiesen.

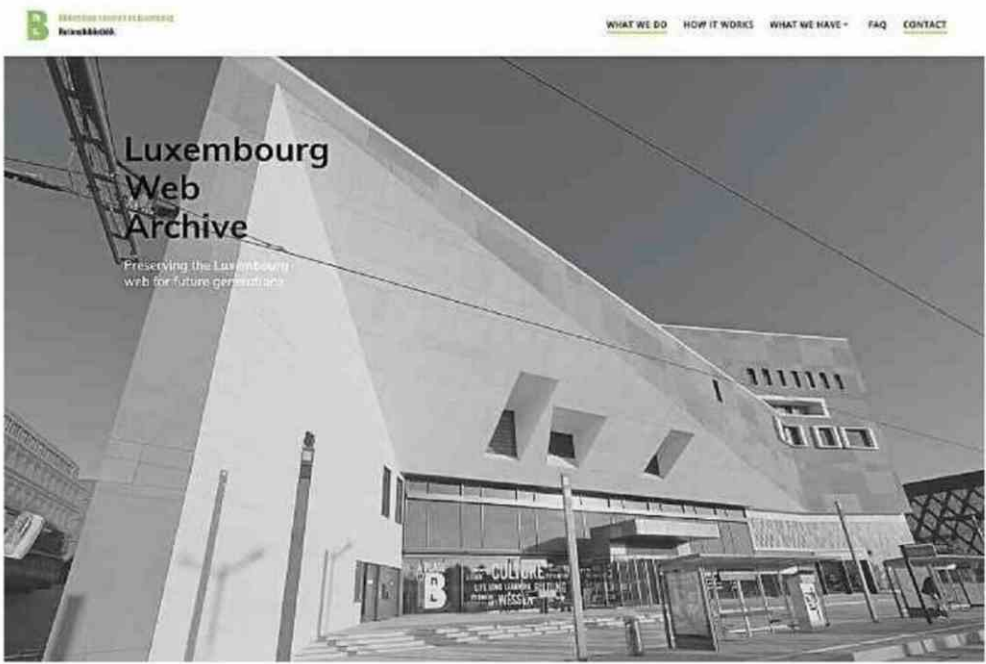
Das Internet ist ein Abbild der verschiedenen Facetten unserer Gesellschaft. Ein besseres Verständnis des Luxemburger Webs impliziert auch ein besseres Verständnis unseres kulturellen Erbes. Ohne die Archivierung des Internets wird das geschichtliche, kulturelle und soziale Gedächtnis der heutigen Zeit in Zukunft riesige Lücken aufweisen. Aus diesem Grund ist das Webarchiv relevant für alle Besitzer von Webseiten, alle Autoren von Texten, Bildern oder Videos – prinzipiell für jeden, der sich online informiert und dieses Privileg auch für zukünftige Generationen sichern möchte.

* Ben Els arbeitet als Digital Curator in der Nationalbibliothek



Warte [Die]

► 17 June 2021



Zur Aufgabe der Nationalbibliothek gehört auch das Sammeln von Web-Inhalten, die im nationalen Webarchiv zusammengetragen und aufbewahrt werden.

