

Webarchivierung in NRW – ein Werkstattbericht

Dr. Andrea Ammendola (ULB Münster)



Ausgangslage

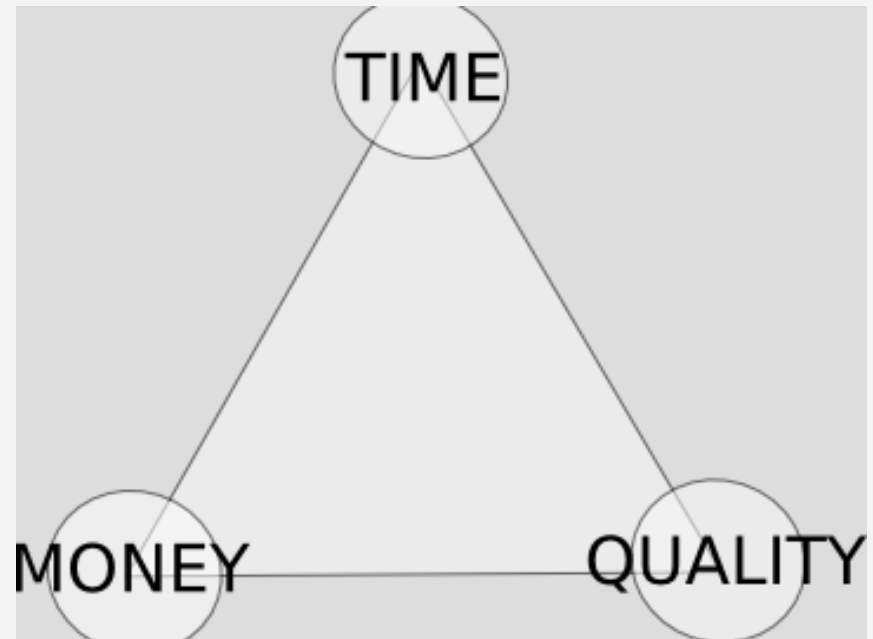
- Novellierung Pflichtexemplargesetz NRW 2013
- Sammlung von elektronischen Publikationen im Rahmen der E-Pflicht in NRW
- Webarchivierung in NRW noch ein Desiderat
- LBB entscheiden gemeinsam (wie bei E-Pflicht und E-Paper)



<https://upload.wikimedia.org/wikipedia/commons/7/76/Emblem-section.svg>

Ziele der LBB

- Kooperation
- Standardisierung
- Ressourcen sparen
- Quantität und Qualität
- Schnelligkeit
- Zugänglichkeit
- Nachhaltigkeit
- Langzeitarchivierung



<https://upload.wikimedia.org/wikipedia/commons/c/c1/Time-Quality-Money.png>

Testphase mit der DNB



- Erste Gespräche im Oktober 2019
- Konzept und Idee der DNB waren überzeugend, insbesondere der Zugriff auf das **gesamte** DNB-Webarchiv für kooperierende Pflichtexemplar-Bibliotheken
- Start einer ersten Testphase ab März 2020
 - Zugang zum DNB-Webarchiv
 - Auswahl von Websites für einen Testcrawl

Testphase mit der DNB



- März–April 2020: Crawl der gemeldeten Testseiten
 - Kinderherzen, Germanwatch, Bonn im Wandel, Histrhen, Kölner Philharmonie, Verband der Ortsvereine Bonn-Oberkassel
 - Geologischer Dienst NRW, NRW-Bank, Henkel, RWE
 - WWU Münster, ULB Münster, LWL, Stadt Dortmund, Bertelsmann AG, Evangelische Kirchen aus Westfalen
 - Nachträglich: Event-Harvesting: Corona in NRW

Testphase mit der DNB



- Mai 2020: Evaluation der Testseiten durch die LBB
 - Erreichbarkeit und Zugriff (Ladegeschwindigkeit und Zugriff teils problematisch)
 - Linktiefe (teilweise fehlen Informationen auf tieferliegenden Ebenen, sehr häufig die Fehlermeldung „404: not found“)
 - Performance (unterschiedliche Ergebnisse: dynamische Elemente werden meist nicht gecrawlt, allerdings auch viele Bilder nicht)

Testphase mit der DNB



- Bisher eingesetztes Personal in den LBB:
 - Pro Einrichtung eine(n) Hauptverantwortliche(n):
 - Auswahl der Seiten für den Testcrawl
 - Evaluation der Testseiten
 - keine genaue Angabe in VZÄ möglich, da die Aufgaben bisher „so nebenbei“ mit erledigt wurden

Testphase mit der DNB



- Juni 2020: Rückmeldung seitens der DNB
- Gründe für die teils heterogenen Ergebnisse:
 - Technische Struktur der Websites sowie dynamische/interaktive Seiteninhalte
 - Stand der vom Dienstleister oia eingesetzten Crawltechnologie
 - Einsatz von Adblockern der Websitebetreiber

Status quo und Ausblick



- Seit Juni 2020: Weitere Gespräche und weitere Tests
- Ziele:
 - Crawlergebnisse verbessern
 - Performance verbessern
 - Workflow mit Qualitätskontrolle entwickeln (inkl. Erschließungsfragen)
 - Produktivphase planen (inkl. Personaleinsatz)
 - Kooperationsvertrag abschließen (inkl. Kostenfrage)