

Herzlich willkommen zum 29. Newsletter des nestor-Kooperationsverbundes!

Wir begrüßen das Konrad-Zuse-Zentrum für Informationstechnik, das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) und das Bundesarchiv bei nestor. Wir berichten vom 4C-Projekt, von zwei Workshops zur Nutzung von Heritrix, und laden Sie dazu ein, an der nestor-school teilzunehmen oder sich der neuen nestor-AG "Formaterkennung" anzuschließen. Außerdem weisen wir auf die Veröffentlichung unseres Jahresberichts 2013 und der Jahresplanung 2014 hin und auf ein Sonderheft der nestor-Edition.

die nestor-Newsletter Redaktion

Neuer nestor Partner I: Bundesarchiv

Das Bundesarchiv hat den gesetzlichen Auftrag, das Archivgut des Bundes auf Dauer zu sichern, nutzbar zu machen und wissenschaftlich zu verwerten. Archiviert werden schriftliche Zeugnisse (bisher ca. 300 Regalkilometer) sowie Bilder, Töne und Filme, die bei zentralen Stellen des Heiligen Römischen Reiches (1495-1806), des Deutschen Bundes (1815-1866), des Deutschen Reiches (1867/71-1945), der Besatzungszonen (1945-1949), der Deutschen Demokratischen Republik (1949-1990) und der Bundesrepublik Deutschland (seit 1949) entstanden sind. Darüber hinaus überliefert das Bundesarchiv auch Nachlässe bedeutender Personen, Unterlagen von Parteien, Verbänden und Vereinen mit überregionaler Bedeutung. Neuerdings nimmt es auch die Aufgaben eines zentralen deutschen Filmarchivs wahr.

Im Bereich der Langzeitarchivierung blickt das Bundesarchiv auf langjährige Erfahrungen im Bereich der Archivierung elektronischer Unterlagen zurück. Die Übernahme "maschinenlesbarer Daten" von Stellen der Deutschen Demokratischen Republik machte in den 1990er Jahren eine vertiefte Beschäftigung mit der dauerhaften Erhaltung und Nutzbarmachung dieser Daten notwendig. Das Bundesarchiv begann 2006 mit dem Aufbau eines Digitalen Archivs. Aktuell stehen die Einführung eines Digitalen Zwischenarchivs des Bundes (Langzeitspeicher für Altdaten der Bundesverwaltung) sowie der Aufbau eines digitalen Geheimarchivs im Zentrum der Aktivitäten.

Tätigkeitsschwerpunkte des Bundesarchivs liegen im Bereich der vorarchivischen Beratung sowie der Bewertung und Übernahme elektronischer Unterlagen. Zurzeit können Unterlagen aus Vorgangsbearbeitungssystemen (E-Akten), Daten aus Filesystemen und Datenbanken übernommen werden.

Der Bestandserhaltung sowie der Überwachung aller Prozesse im Digitalen Archiv und der Pflege relevanter Informationen dient ein auf diese Aufgabe zugeschnittenes Tool des Bundesarchivs.

Neuer nestor-Partner II: Hochschulbibliothekszentrum (hbz) NRW

Das hbz verfolgt mit seinem Aufgabenbereich Langzeitarchivierung einen ganzheitlichen Ansatz, der neben den gängigen technischen Aspekten insbesondere organisatorische und strategische Konzeptionsarbeiten einschließt. Seit vielen Jahren beteiligt sich das hbz deshalb an unterschiedlichen nestor Arbeitsgruppen.

Das hbz berät, ermittelt und deckt den Bedarf an modularen Lösungen für die Langzeitarchivierung. Es gestaltet die Infrastruktur für eine Langzeitverfügbarkeit von digitalen Daten sowohl auf Landes- als auch auf Bundesebene aktiv mit. So ist das hbz Partner im Projekt "Digitales Archiv NRW" (DA NRW) zum Aufbau einer Kultursparten übergreifenden Langzeitverfügbarkeitslösung des Landes Nordrhein-Westfalen. Es übernimmt hier als Infrastrukturpartner insbesondere Aufgaben für die Langzeitarchivierung und -verfügbarkeit von Materialien aus Universitäts- und Landesbibliotheken.

Mit dem System Edoweb betreibt es außerdem eine umfassende Lösung für die Archivierung des elektronischen Pflichtexemplars beim Landesbibliothekszentrum Rheinland-Pfalz.

Die Teilnahme am deutschen LOCKSS-Netzwerk und das Angebot modularer Services wie der PDF/A-Konvertierung ermöglichen darüber hinaus die Umsetzung bedarfsorientierter Lösungen für unterschiedliche Kundengruppen.

Aktuell engagiert sich das hbz in den nestor AGs "Kooperative Langzeitarchivierung", "Kosten" sowie "Policy-Entwicklung" und bringt dort auch eigene Forschungs- und Entwicklungsarbeiten in das Kompetenznetzwerk ein.

Neuer nestor-Partner III: Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

Das Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) ist ein außeruniversitäres Forschungsinstitut für angewandte Mathematik und Informatik

in Trägerschaft des Landes Berlin. Das ZIB betreibt zusammen mit der Universität Hannover den Supercomputer HLRN III für die norddeutschen Bundesländer sowie das Berliner Wissenschaftsnetz BRAIN und ist ein Betreiber des Forschungscampus MODAL. In der Abteilung Wissenschaftliche Information des ZIB arbeiten drei Einrichtungen und Arbeitsgruppen an Lösungen für die Datenanalyse und Datenerschließung, für die Online-Präsentation und die technische und semantische Langzeitverfügbarkeit von digitalen Objekten: Der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV), die AG Museumssoftware und die Servicestelle Digitalisierung des Landes Berlin (digiS). Der KOBV ist Dienstleister für wissenschaftliche und öffentliche Bibliotheken und betreibt Entwicklung in den Bereichen Repositorymanagement, Forschungsdatenmanagement und Discoverysysteme. Die AG Museumssoftware entwickelt und betreut die Dokumentationssoftware GOS, die zur Erschließung und Verwaltung von musealen Sammlungen und Beständen eingesetzt wird und forscht im Bereich Datenmodellierung und Datenaustausch. Die Servicestelle Digitalisierung koordiniert und berät Teilnehmer am Förderprogramm der Berliner Kulturverwaltung zur Digitalisierung von Kulturgut aus Museen, Archiven, Bibliotheken und Gedenkstätten. Um die gemeinsame Aufgabe der Sicherstellung der Langzeitverfügbarkeit von digitalen Objekten umsetzen zu können, wird derzeit ein Langzeitarchiv auf Basis der Open-Source-Software Archivematica am ZIB aufgebaut.

Aufruf zur Mitarbeit in der neuen nestor-AG "Formaterkennung"

Formatidentifizierung, Charakterisierung und Validierung bei der Übernahme von Daten in das Langzeitarchiv sind wichtige Grundvoraussetzungen für die Risikoabschätzung und spätere strategische Entscheidungen wie Migration und Emulation. Bereits vorhandene Werkzeuge wie JHOVE, DROID, FITS weisen aufgrund der Fülle der vorhandenen Formate große Lücken auf. Auch internationale Nachweissysteme wie die Format Registries PRONOM oder UDFR müssen daher regelmäßig ergänzt werden.

Ziel der AG ist die Zuarbeit zu kooperativ betriebenen Nachweissystemen wie PRONOM und die Auseinandersetzung mit Werkzeugen wie JHOVE und DROID.

Wir möchten nicht nur vorhandene Lücken identifizieren, sondern auch einen Beitrag dazu leisten, diese systematisch zu schließen und somit das Identifizieren und Validieren der Formate für die gesamte Community vorantreiben.

Wir suchen Kolleginnen und Kollegen für die Mitarbeit in der neuen nestor-AG "Formaterkennung". Sofern Sie daran interessiert sind, gemeinsam mit Fachkolleginnen und -kollegen zum Thema beizutragen, sind Sie herzlich eingeladen, sich zu beteiligen.

Melden Sie sich bitte bis zum 31.03. bei Yvonne Friese (y.friese@zbw.eu).

Fortschritte in der Heritrix-Nutzung

Heritrix etabliert sich im Bereich der Webseitenarchivierung als Crawler der ersten Wahl: laufende Entwicklung, Möglichkeiten zur Parametrisierung, das selbst beschreibende Speicher- und Archivformat WARC sowie die Verbindung zur Wayback Machine sind Funktionen, die die Vorrangstellung begründen. In dieser Vielfalt liegen aber auch die Schwierigkeiten der Heritrix-Nutzung und -Integration: die Dokumentation der Parametrisierung ist auf viele Stellen verteilt und zum Teil unzureichend. Wer versteht z.B., was eine TooManyHopsDecideRule ist? Und wer übersieht, welche Auswirkungen das Setzen einer solchen Rule in umfangreichen Harvestingläufen hat?

Heritrix wird meist nicht eigenständig verwendet, sondern ist Teil umfassender Workflowtools wie z.B. dem WebCurator Tool, der NetarchiveSuite oder auch [SWBcontent](#), das seit 2002 beim BSZ entwickelt und in etlichen Installationen betrieben wird.

Im Rahmen von zwei Expertenworkshops fand ein intensiver Austausch über die breiten Nutzungsmöglichkeiten von Heritrix statt:

Auf Einladung des BSZ traf sich der Anwenderkreis von SWBcontent am 3. Dezember 2013 zu einem Workshop, der die Erläuterung der grundsätzlichen Arbeitsweise des Crawlers, die Implementation und Parametrisierung sowie die Logauswertung mit praktischen Übungen verband. Dieser Workshop wurde nötig, nachdem Heritrix mit einer erweiterbaren Auswahl von Parametrisierungsmerkmalen in die Ablaufsteuerung und das User Interface von SWBcontent integriert worden war.

Bereits eine Woche später fand auf Einladung der BSB am 10. Dezember 2013 ein zweiter Workshop für aktive Heritrix-Nutzer im deutschsprachigen Raum statt. Die Teilnehmenden kamen vom BSZ, von der Österreichischen und der Schweizerischen Nationalbibliothek sowie der BSB selbst. Ziel des

Workshops war, praxisnah und konkret die Konfigurationsmöglichkeiten und -grenzen von Heritrix zu analysieren und sich über etwaige Lösungswege auszutauschen. Trotz unterschiedlicher Webarchivierungs-Tools und verschiedener Heritrix-Versionen, stehen alle Anwender vor ähnlichen Herausforderungen: z.B. stellt das Crawlen von Javascript- oder Flash-Anwendungen sowie von Websites, die sich aus "Frames" zusammensetzen, für alle eine kaum lösbare Schwierigkeit dar.

Es bestand Einigkeit darüber, dass eine detaillierte Qualitätskontrolle gerade beim selektiven Crawlen wichtig ist und auf keinen Fall allein auf einer visuellen Überprüfung basieren darf. Gängige Präsentationstools hinken der Entwicklung hinterher und können Inhalte oft nicht korrekt darstellen, auch wenn sie von Heritrix korrekt gecrawlt wurden und in der WARC-Datei vorhanden sind.

Die Teilnehmenden begrüßten einen weiteren konkreten Austausch auf operativer Ebene, der zunächst virtuell weitergeführt wird.

4C –Projekt geht in die Halbzeit

Das 4C Projekt – "Collaboration to Clarify the Cost of Curation" ist ein Konsortium von 13 Partnern in 7 Ländern, welches mit seiner Arbeit Organisationen in ganz Europa dabei unterstützt, durch enge Zusammenarbeit ein besseres Verständnis über Kosten in der digitalen Bestandserhaltung zu gewinnen und so effektiver in die Langzeitarchivierung investieren zu können. Aktuell hat das Projekt seine "Halbzeit" erreicht; alle Informationen, Hintergründe und Neuigkeiten dazu finden Sie [hier](#). Im neuesten 4C-Blog können Sie eine [Bestandsaufnahme](#) des Projektkoordinators Neil Grindley (Jisc) nachlesen und kommentieren. Der 4C-Blog ist für Gastbeiträge zum Thema Kosten in der digitalen Langzeitarchivierung oder verwandten Themen offen und lädt zur Beteiligung ein. Das 4C-Team erreichen Sie [hier](#) oder über Katarina Haage (k.haage@dnb.de).

Einladung zur nestor School 2014

Das nestor-Team freut sich, zur nestor School 2014 einladen zu können, die sich dem Themenkomplex "Digitalisierung und Langzeitarchivierung" widmet. Die mehrtägige Fortbildungsveranstaltung wird vom 3. bis 5. März 2014 im [Cyriakushaus in Gernrode/Harz](#) stattfinden, durchgeführt von nestor und Hochschulpartnern, die sich auf der Grundlage eines [Memorandum of Understanding](#) zum Ziel gesetzt

haben, die Qualifizierung auf dem Gebiet der digitalen Langzeitarchivierung voranzutreiben. Zudem wird die School in diesem Jahr vom EU-Projekt [DASISH](#) unterstützt.

Teilnehmerinnen und Teilnehmer erwartet ein interessantes Programm, das sich u.a. mit Fragen der Auswahl und Priorisierung, der Scanqualität, geeigneten Formaten sowie Workflows und Metadaten im Hinblick auf die Langzeitarchivierung beschäftigen wird. Die einzelnen Sessions werden von Experten vorgestellt und anschließend in praxisnahen Übungen vertieft. Die Veranstaltung bietet Gelegenheit zum intensiven Austausch mit Kolleginnen und Kollegen sowie zum Knüpfen von Kontakten und dem Auf- und Ausbau von Netzwerken.

Eine Webseite mit Details zum Programm, weiteren Informationen und der Möglichkeit sich anzumelden finden Sie [hier](#).

nestor-Bericht 2013 und Planung 2014

Der Tätigkeitsbericht des Kompetenznetzwerkes für 2013 und die Jahresplanung für 2014 sind [hier](#) online einsehbar.

Sonderheft der nestor-edition

Im Rahmen der nestor-edition sind die Beiträge eines Workshops zur digitalen Langzeitarchivierung auf der Informatik 2013 erschienen. Darin enthalten:

- J. Enge, H. W. Kramski und T. Lurk: Ordnungsstrukturen von der Floppy zur Festplatte. Zur Vereinnahmung komplexer digitaler Datensammlungen im Archivkontext
- W. Zenk-Möltgen und M. Linne: datorium – ein neuer Service für Archivierung und Zugang zu sozialwissenschaftlichen Forschungsdaten
- A. Weisser: Digitale Langzeitarchivierung von Videokunst
- S. Cuy, M. Fischer, D. de Oliveira, J. Peters, J. Puhl, L. Rau und M. Thaller: DA-NRW: Eine verteilte Architektur für die digitale Langzeitarchivierung
- D. von Suchodoletz, K. Rechert, I. Valizada und A. Strauch: Emulation as an Alternative Preservation Strategy - Use-Cases, Tools and Lessons Learned
- H. Kalb, P. Lazaridou, V. Banos, N. Kasioumis und M. Trier: BlogForever: From Web Archiving to Blog Archiving
- S. Schwalm, U. Korte und D. Hühnlein: Vertrauenswürdige und beweiswerterhaltende elektronische Langzeitspeicherung auf Basis von DIN 31647 und BSI-TR-03125

Die Veröffentlichung findet sich [hier](#).

Termine

3. bis 5. März 2014 [nestor Spring-School](#) "Digitalisierung und Langzeitarchivierung" in Gernrode/Harz

11. und 12. März 2014 [18. Treffen des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen"](#) in Weimar

2. April 2014 [Öffentlicher Workshop der nestor AG Policy](#) zur Erstellung institutioneller Polycys

13. bis 16. Mai 2014 [IS&T Archiving Conference](#) in Berlin

26. Juni 2014 [nestor-Praktikertag 2014](#) zum Thema "Webarchivierung" in Hannover

nestor-Mailingliste

Möchten Sie an der Diskussion um digitale Langzeitarchivierung teilnehmen?

Dann abonnieren Sie sich bitte für die nestor-Mailingliste unter:

www.langzeitarchivierung.de/services/maillingliste/maillingliste.htm

Diese Liste ist offen für alle Fragen zu den Themen digitale Langzeitarchivierung und nestor.

nestor-Newsletter

Der Newsletter wird von den nestor-Partnern kooperativ erstellt. Bei inhaltlichen Fragen, Anregungen, Kommentaren zu dem Newsletter wenden Sie sich bitte an Armin Straube: a.straube@dnb.de

Zum Bestellen schicken Sie bitte eine E-Mail an lists@langzeitarchivierung.de mit dem Betreff: subscribe langzeitarchivierung-newsletter.

Zum Abmelden schicken Sie bitte eine E-Mail an lists@langzeitarchivierung.de mit dem Betreff: unsubscribe langzeitarchivierung-newsletter.

nestor-Geschäftsstelle

Deutsche Nationalbibliothek
Adickesallee 1
60322 Frankfurt am Main
Tel.: +49-69-1525-1141
Fax: +49-69-1525-1799
E-Mail: VL-nestor@dnb.de
www.langzeitarchivierung.de