

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 12.3

Technisches Workflowmanagement
in der Praxis:
Erfahrungen und Ergebnisse

nestor 

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Neben der Online Version 2.3 ist eine Printversion 2.0 beim Verlag Werner Hülsbusch,
Boizenburg erschienen.

Die digitale Version 2.3 steht unter folgender Creative-Commons-Lizenz:
„Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 3.0
Deutschland“
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

URL für Kapitel 12.3 „Technisches Workflowmanagement in der Praxis: Erfahrungen und
Ergebnisse“ (Version 2.3): [urn:nbn:de:0008-20100617256](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617256)
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100617256>



Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.

12.3 Technisches Workflowmanagement in der Praxis: Erfahrungen und Ergebnisse

Reinhard Altenböner

Massenprozesse in der Langzeitarchivierung sind noch wenig etabliert; daher ist wie bereits festgestellt der Umfang praktischer Erfahrungen noch begrenzt. Wichtige Erkenntnisse konnte sowohl in der technischen Workflowentwicklung als auch in der praktischen Umsetzung die niederländische Nationalbibliothek sammeln. Auch in der Deutschen Nationalbibliothek liegen erste Erfahrungen vor¹⁵: Nach einer Gesetzesnovelle Mitte des Jahres 2006 hat sie die Zuständigkeit für die Erhaltung der Langzeitverfügbarkeit deutscher Online – oder Netzpublikationen erhalten und steht nun vor sehr konkreten Herausforderungen, die derzeit zu einer umfassenden Reorganisation des technischen Workflow führen.¹⁶ Mit dem Inkrafttreten des neuen Gesetzes und der damit verbundenen deutlich erweiterten Verpflichtung, die Aufgabe der Langzeitarchivierung zu erfüllen, stellt sich hier die Frage in einer neuen Dimension: Wie wird die Bibliothek die neuen Abläufe organisieren, welche technischen Methoden und Anwendungen werden im Massenverfahren eingesetzt? Da gleichzeitig die alten Arbeitsabläufe und –verfahren weiterlaufen, stellt sich die Frage der Integration. Zwar ist die Bibliothek in der glücklichen Situation, für die neuen Aufgaben zusätzliche Ressourcen erhalten zu haben, doch würden diese nicht eine nahtlose Imitation des organisatorisch-operativen Workflows auf Basis der existierenden Systeme abdecken – das ergibt sich schon aus den Mengen, um die es geht.

Die Königliche Bibliothek der Niederlande (KB) betreibt seit dem Jahr 2003 das OAIS-kompatible Archivierungssystem DIAS der Firma IBM operativ und hat im Laufe der gewonnenen Erfahrungen insbesondere organisatorisch eine ganze Reihe von Anpassungen vorgenommen.¹⁷ Technisch gesehen wurde

15 Vgl. hierzu den einführenden Artikel von Maren Brodersen / Sabine Schrimpf im 18. Kapitel „Praxisbeispiele“ dieser Enzyklopädie unter dem Titel „Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek“:

16 Es sei angemerkt, dass es eine ganze Reihe von weiteren Publikationen zum Thema gibt. So stellte etwa Clifton (2005) Workflows der australischen Nationalbibliothek vor; diese beziehen sich allerdings auf die manuelle Behandlung von Objekten mittels einzelner Tools. Seit 2007 läuft in der australischen Nationalbibliothek ein Projekt zur Etablierung und IT-basierter Unterstützung der Datenmigration von physischen Datenträgern; noch ist es zu früh, um die Übertragbarkeit bzw. Nachnutzbarkeit des Ansatzes beurteilen zu können, vgl. <http://prometheus-digi.sourceforge.net/>

17 KB (2008)

eine auch in der KB weitgehend isolierte eigene Entwicklung aufgesetzt, die nur in geringem Maße an die sonstigen Abläufe der Bibliothek angebunden ist. Schwerpunkt liegt auf dem Ingest-Prozess, also dem Einspielen des in der Regel von Verlagen bereitgestellten publizierten Materials in das Archiv. Dieses erfolgt weitgehend automatisiert und es ist der Niederländischen Nationalbibliothek sehr schnell gelungen, die Fehlerquoten auf niedrige Promillebereiche zu drücken. Inzwischen sind mehr als siebzehn Millionen Objekte eingespielt, darunter auch (allerdings wenige) komplexe Objekte wie historische CD-ROMs. Für alle Objekte – es handelt sich in der weit überwiegenden Zahl um PDF-Dateien – gilt, dass in der eigentlichen Langzeitarchivumgebung nur rudimentäre Metadateninformationen gespeichert werden; die bibliographischen Informationen werden über das Recherchesystem der KB zur Verfügung gestellt.

Insgesamt ist es der KB gelungen, den technischen Workflow relativ unkompliziert und damit effizient und für hohe Durchsatzmengen geeignet zu gestalten. Dies war auch deswegen möglich, weil die Zahl der Lieferanten in das System in den Niederlanden zumindest in der Startsituation klein war, da wenige große Verlage einen überwiegenden Anteil am Publikationsvolumen der Niederlande haben.

In Deutschland stellt sich die Situation anders dar: Hier bestimmen in einer zum Teil noch sehr traditionell geprägten Veröffentlichungslandschaft viele Verleger das Bild. Ausgangspunkt für die Deutsche Nationalbibliothek bei der Neukonzipierung ihrer technischen Abläufe war eine Situation, in der für die Verarbeitung von Online-Dokumenten bereits eine Vielzahl von mehr oder weniger halbautomatischen Verfahren für Netzpublikationen, Online-Dissertationen und weitere Materialien existierte. Diese historisch gewachsenen Strukturen standen nebeneinander, d.h. – nicht untypisch für Gedächtnisorganisationen im öffentlichen Kontext – die einzelne Objektklasse war der definitorische Ausgangspunkt für einen hochspezialisierten Workflow. Ziel war und ist daher die Schaffung eines automatischen, einheitlichen Verfahrens mit der Übergabe der Archivobjekte an das im Rahmen des Projekts kopal entstandene Archivsystem und die dort entstandenen Verfahren.¹⁸ Davon betroffen sind sowohl der Ingest wie aber auch der Zugriff auf die Objekte: Aus der Langzeitarchivlösung kopal werden Objekte an den Arbeitsplatzrechner übergeben oder über das in der Realisierungsphase befindliche Bereitstellungssystem zur Verfügung gestellt. Dabei sind zahlreiche Arbeitsbereiche in der DNB involviert: neben dem bibliographischen System sind dies die Fachbereiche, externe Ablieferer,

18 kopal (2008)

aber auch die für die digitalen Dienste der DNB Verantwortlichen. Insofern ist hier vieles noch offen und ein Werkstattbericht mag dies illustrieren:¹⁹

Für den Transfer und das Angebot von Objekten auf elektronischen Materialien auf physischen Datenträgern (d.h. CD- bzw. DVD-Veröffentlichungen) existiert ein älterer, segmentierter Workflow, der nun aufgrund der Anforderungen seitens Archivsystem und künftiger Bereitstellung anzupassen ist. Hierfür wurde eine kommerzielle Lösung der Fa. H+H ausgewählt und an die spezifischen Erfordernisse der DNB angepasst. Nach Erstellung der Images der Daten auf Anforderung hin werden die Daten dem Benutzer zur Verfügung gestellt und zwar in der betriebstechnischen Ausgangsumgebung (zum Beispiel Windows 95), für die sie einmal erstellt wurden. Allerdings ist dieser Prozess Ad-Hoc-Bereitstellung noch nicht an die Langzeitarchivierung der DNB angebunden. Der systematische Transfer von Daten auf physischen Trägern bezieht sich in der DNB aktuell vor allem auf Audio-Dateien, für die im Rahmen eines Vorprojekts ein Workflow entwickelt wurde, der neben dem Rippen der CDs selbst auch das Scannen aller Begleitmaterialien beinhaltet. Die Verknüpfung mit vorhandenen Metadaten und die Anreicherung mit weiteren Informationen aus externen Quellen sind weitere inhaltliche Elemente des Vorgehens. Kennzeichnend für den Workflow ist insbesondere die besondere Bedeutung der Qualitätssicherung einerseits sowie die Dokumentation der Ergebnisse der Transferläufe andererseits. Weitergehende Aspekte beziehen sich auf die Skalierung, um das Ziel von 500 migrierten CD pro 24 Stunden zu erreichen.

Für genuin online vorliegende Netzpublikationen wurde der Workflow unter Einbeziehung der Anforderungen der Langzeitarchivierung neu gestaltet und auf die Schnittstellen des Archivsystems angepasst. Dabei ergeben sich eine ganze Reihe von Problemen: So entsprechen fortlaufende Publikationen (vor allem elektronische Zeitschriften-Artikel) und die künftigen zu archivierenden Objekte häufig nicht der aktuellen Abbildung im Online-Katalog. Bibliografische Metadaten von Archivobjekten müssen aber künftig im bibliografischen System abgebildet werden, um einen einheitlichen Zugang zu gewährleisten. Dazu müssen eine Festlegung von Erschließungsvarianten und ein Mapping von Archivobjekten auf Katalogobjekte erfolgen, letztlich also eine klare Definition der Granularität von Objekten und ihrer Abbildung gefunden werden.

Das URN-Management in der DNB wurde bereits erweitert und vor allem technisch so weiterentwickelt, dass eine Einbindung in andere Arbeitszusammenhänge/Module erfolgen kann. Da jedes Objekt zum Einspielen in das Ar-

19 Wollschläger (2007), S. 18ff.

chiv einen Persistent Identifier benötigt, erfolgt für bereits gesammelte Objekte ohne URN eine retrospektive Vergabe der URN. Alle neuen Objekte müssen entweder mit URN geliefert werden bzw. bei Eingang/Bearbeitung einen URN erhalten, was dem künftigen Verfahren entspricht.

Wesentliche Voraussetzungen für die Einbindung des Archivs in die Geschäftsumgebung der Institution liegen mittlerweile vor oder werden gerade geschaffen. Insbesondere die Kernelemente des Produktionssystems laufen, das produktive Einspielen von Material wurde und wird erprobt, nötige Weiterentwicklungen (z.B. noch fehlende Module zur Auswertung von Dateiformaten) wurden und werden ermittelt und Änderungen / Anpassungen in diversen Workflows der traditionellen Bearbeitung wurden bereits angestoßen. Weitere Aufgaben betreffen in hohem Maße die Übergabe des kopal-Systems, die Etablierung einer ständigen Arbeitseinheit sowie die retrospektive Aufarbeitung des früher bereits in die Bibliothek gelangten Materials.

Hinter diesen Bemühungen steht der Anspruch, die neuen, mit der Gesetzesnovelle übernommenen Aufgaben, die weit über das Arbeitsfeld der Langzeitarchivierung hinausgehen, in einem ganzheitlichen technischen Workflow abzubilden. In dessen Mittelpunkt stehen aktuell die Übernahme von elektronischen Objekten mit möglichst breiter Nachnutzung vorhandener Metainformationen und die Integration der Abläufe in die Arbeitsumgebung der DNB.

Die praktischen Erfahrungen an der DNB insbesondere für diesen Bereich belegen den besonderen Bedarf für eine bewusste Modellierung der Geschäftsprozesse, die in der Vergangenheit häufig nur unvollkommen gelungen ist. Im Ergebnis standen isolierte, von nur wenigen Personen bediente und bedienbare Abläufe mit einem hohen manuellen Eingriffs- und Fehlerbehandlungsbedarf. Ohne dass heute bereits ein komplettes Profil der zukünftigen technischen Workflow-Umgebung vorliegt, kann doch gesagt werden, dass ein methodisch bewusstes, in enger Kooperation von Bedarfsträger und Informationstechnik ablaufendes Vorgehen zu deutlich klareren Vorstellungen darüber führt, wie die wesentlichen Arbeitsschritte exakt aussehen und wie sie adäquat so abgebildet werden, dass die entstehenden Services auch langfristig und damit über ihren aktuellen Entstehungshintergrund hinaus genutzt werden.

Dass dabei für eine technische Arbeitsumgebung besondere Anforderungen an die Flexibilität und die Orientierung an offenen Standards gelten, liegt auf der Hand und hat wesentlich die Entwicklungsleitlinien für kopal mitbestimmt.²⁰

20 kopal (2008a)

Quellenangaben

- Clifton, Gerard: Safe Havens In A Choppy Sea: Digital Object Management Workflows At The National Library of Australia (2005), Beitrag zur iPRES - International Conference on Preservation of Digital Objects, Göttingen (September 15, 2005). In: [http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Safe Havens In A Choppy Sea Digital Object Management Workflows At The National Library of Australia - Gerard Clifton.pdf](http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Safe%20Havens%20In%20A%20Choppy%20Sea%20Digital%20Object%20Management%20Workflows%20At%20The%20National%20Library%20of%20Australia%20-%20Gerard%20Clifton.pdf) (Zugriff 15.2.2010)
- Keller, Gerhard / Nüttgens, Markus / Scheer, August-Wilhelm (1992): *Semantische Prozessmodellierung auf der Grundlage „Ereignisgesteuerter Prozessketten (EPK)*. In: A.-W. Scheer (Hrsg.): Veröffentlichungen des Instituts für Wirtschaftsinformatik, Heft 89, Saarbrücken. Online in: <http://www.iwi.uni-sb.de/Download/iwihefte/heft89.pdf> (Zugriff 15.2.2010)
- Königliche Bibliothek der Niederlande (KB): The e-Depot system (DIAS) (2010) In: <http://www.kb.nl/dnp/e-depot/operational/background/index-en.html> (Zugriff 15.2.2010)
- Kopal (2008): Projekthompae. In: <http://kopal.langzeitarchivierung.de/> (Zugriff am 15.2.2010)
- Kopal (2008a): *kopal: Ein Service für die Langzeitarchivierung digitaler Informationen*. In: http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf (Zugriff am 15.2.2010)
- Martin, Norbert (1999): *Und wie kommt die Dissertation auf den Server? Gedanken zum Workflow. Vortrag auf der IuK-Tagung „Dynamic Documents“, vom 22.-24.3.1999 in Jena*. In: <http://edoc.hu-berlin.de/epdiss/jena3/workflow.pdf> (nicht mehr über diesen Server erreichbar, leider auch nicht im Internet Archive gespeichert (15.2.2010)
- Object Management Group: Unified Modeling Language (UML), version 2.2. In: <http://www.omg.org/technology/documents/formal/uml.htm> (Zugriff am 15.2.2010)
- Stapel, Johan: *The KB e-Depot. Workflow Management in an Operational Archiving Environment* (2005). Beitrag zur iPRES - International Conference on Preservation of Digital Objects, Göttingen (September 15, 2005). In: <http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Workflow%20Management%20In%20An%20Operational%20Archiving%20Environment%20-%20Johan%20Stapel.pdf> (Zugriff 15.2.2010)
- Verwaltungslexikon (2008) *Management und Reform der öffentlichen Verwaltung* (2008) In: <http://www.olev.de/w.htm> (Zugriff am 15.2.2010)

Wollschläger, Thomas (2007): „*kopal goes live*“. In: Dialog mit Bibliotheken 19 (2007), H.2, S. 17 – 22

Workflow Management Coalition (2008) – Website. In: <http://www.wfmc.org/> (Zugriff am 15.2.2010)