

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

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 12.2

Workflow in der Langzeitarchivierung:
Methode und Herangehensweise

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 12.2 „Workflow in der Langzeitarchivierung: Methode und Herangehensweise“
(Version 2.3): <urn:nbn:de:0008-20100305233>
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100305233>



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

12.2 Workflow in der Langzeitarchivierung: Methode und Herangehensweise

Reinhard Altenböner

Die allmähliche Einführung der Langzeitarchivierung in das reguläre Auftragsportfolio von Bibliotheken und anderen Kulturerbeeinrichtungen mit immer höheren Bindungsquoten von Personal und anderen Ressourcen erzeugt(e) zunächst neue, häufig isolierte und händisch durchgeführte Abläufe. In ganzheitlichen Betrachtung aber verändern sich Arbeitsabläufe und die sie modellierenden Geschäftsprozesse. So ist schon für sich genommen die Einspielung von Daten in ein Langzeitarchiv ein komplexer Vorgang, in dem eine ganze Reihe von auf einander bezogenen bzw. von einander abhängenden Aktivitäten ablaufen. Vor allem aber die zunehmende Relevanz der technischen und operativen Bewältigung der Aufgabe verlangt nach einer systematischen Modellierung der Geschäftsprozesse, also dem Einstieg in ein systematisches (technisches) Workflowmanagement. Es gilt allerdings festzustellen, dass selbst in Einrichtungen, die bereits seit einigen Jahren Erfahrungen mit dem Betrieb von Langzeitarchiven und ihrer Integration in die jeweilige Systemlandschaft gesammelt haben, häufig noch isolierte Bearbeitungsketten ablaufen, die zudem keinesfalls wirklichen Vollständigkeitsgrad haben, also alle Anforderungs- / arbeitsfelder abdecken und außerdem vielfach noch manuelle Eingriffe erfordern, insbesondere auf dem Gebiet des Fehlermanagements. In aller Regel sind diese Abläufe nicht massenfähig, d.h. es bestehen Zweifel, ob hohe Volumina transparent prozessiert werden können.⁷

Diese Feststellung bedeutet aber auch, dass der Erfahrungshorizont zum technischen Workflow und insbesondere zum Management insgesamt noch gering ist, also hier noch konkrete Erfahrungen vor allem im Umgang mit großen Mengen und insbesondere auch im automatisierten Qualitätsmanagement gewonnen werden müssen. Insofern hat die Beschäftigung mit dem technischen Workflow derzeit noch stark theoretischen, sozusagen ‚propädeutischen‘ Charakter.

7 Initiativen, die hier bereits erfolgreich agieren, sind die Nationalbibliothek der Niederlande (siehe dazu weitere Information im Abschnitt 14.3) sowie der non-for-profit-Service portico; beide haben diese Situation durch eine konsequente Beschränkung auf bestimmte Objekttypen erreicht. Es handelt sich jeweils um dedizierte Langzeitarchiv-Dienste mit geringem Integrationsgrad in sonst vorhandene Abläufe. Demgegenüber ist LOCKKS (siehe an anderer Stelle in diesem Band) zwar gut in die Abläufe der es tragenden Einrichtungen integriert, allerdings betreibt dieser Dienst im Wesentlichen bit-stream preservation.

In einer Situation, in der verschiedene (bereits existente und neu entwickelte) Arbeitsprozesse ineinander greifen und auch verschiedene Organisationseinheiten an ein und demselben Vorgang beteiligt sind, ist die Modellbildung im Sinne der Geschäftsprozessmodellierung ein Beitrag zu einer umfassenden Optimierung. Damit befinden sich Bibliotheken, Archive und Museen in einer Situation, die man mit den Anstrengungen der Privatwirtschaft Anfang der 1990er Jahre vergleichen kann, als dort die Modellierung von Geschäftsprozessen unter verschärften Wettbewerbs- und Kostendruckbedingungen systematischer als zuvor angegangen wurde. Auch wenn im öffentlich finanzierten Umfeld in besonderem Maße historisch geprägte Organisationsformen gegeben sind, die eine vorgangsbezogene Sicht erschweren, führt an der grundsätzlichen Anforderung der Neu-Modellierung aus systematischer Sicht kein Weg vorbei. Diese wird im Umfeld des technischen Workflow immer stark auch von der informationstechnischen Entwicklungsseite getrieben sein, denn Ziel der Geschäftsprozessmodellierung ist letztlich ihre technische Abbildung.

Übergeordnete Ziele dieses Herangehens, also der systematischen Modellierung und eines methodenbewussten Workflowmanagements und zugleich auch Chance sind⁸:

- Verbesserung der Prozessqualität
- Vereinheitlichung der Prozesse
- Schnellere und zuverlässigere Bearbeitung von Aufträgen (extern und intern)
- Reduzierung der Durchlaufzeiten
- Kostenreduktion
- Verbesserte Verfügbarkeit von Information / Dokumentation
- Erhöhte Prozessflexibilität
- Erhöhung der Transparenz der Prozesse (Statusermittlung, Dokumentation von Entscheidungen), Qualitätssicherung
- Automatische Eingriffsmöglichkeiten: Dokumentation, Eskalation bei Zeitüberschreitungen, Verteilung von Aufgaben und Verantwortlichkeiten
- Vermeidung von Redundanz, mangelnder Aktualität und Inkonsistenz durch Mehrfachschritte

Natürlich lassen sich kleine isolierte Prozesse oder Prozesselemente durch individuelle Programmierung jeweils neu umsetzen. Dies geschah in der Ver-

8 Die folgende summarische Zusammenstellung betrifft sowohl organisatorische wie technische Aspekte des Workflowmanagements. Eine Trennung ist theoretisch zwar möglich, praktisch aber nicht sinnvoll.

gangenheit vielfach für einzelne Objektklassen oder auch einzelne Datenübergabe- oder -tauschprozesse. Aber schon beim Zusammenführen bzw. Hintereinandersetzen der einzelnen Teilschritte bedarf es einer Gesamtlogik für das Management des Ablaufs dieser Schritte. Fehlt diese Logik, entstehen letztlich viele immer wieder manuelle neu anzustößende Teilkonstrukte mit dazu häufig proprietären „Konstruktions“elementen. Schon insofern ist die systematische Analyse verschiedener wiederkehrender Arbeitsabläufe ein sinnvoller Ansatz, um so zur Modellierung auch komplexer Vorgänge im Bereich der Langzeitarchivierung zu kommen.

Ziel dieses systematischen Ansatzes ist es, Services zu definieren, die auch in anderen Kontexten (wieder) verwendbar sind. Sie bilden Arbeitsschritte granular ab, die so in verschiedenen Umfeldern vorkommen (können), beispielsweise das Aufmachen eines Bearbeitungsfalls für ein Objekt und die IT-gestützte Verwaltung verschiedener Be-/Verarbeitungsschritte dieses Objekts. In dieser Perspektive entsteht der Geschäftsprozess für eine Klasse von Objekten aus der Zusammenfügung verschiedener Basisservices, die miteinander interoperabel sind. Dass diese Herangehensweise sehr stark mit dem Modell der Serviceorientierten Architektur (SOA) bei der Entwicklung IT-basierter Lösungen korrespondiert, ist dabei kein Zufall. Voraussetzung dafür ist aber die systematische Modellierung der Arbeits- oder Geschäftsprozesse, die vorgeben, welche Services wann und wie gebraucht werden. Die Prozessmodellierung bildet also die Basis für die Implementierung, die Prozesse selbst dienen der Orchestrierung, dem Zusammenspiel und der Aufeinanderabstimmung der Services. In einem optimalen (Infrastruktur)-Umfeld können so die Arbeitsschritte als kleinere Einheit eines Geschäftsprozesses verschiedene Services lose zusammenbringen.

Der Ansatz, Services nachnutzbar zu gestalten, bezieht sich in der Regel auf eine Organisation. Zwar wird immer wieder versucht, Geschäftsprozesse aus einem institutionellen Umfeld auf ein anderes zu übertragen, allerdings erweist sich dies in der Praxis als außerordentlich schwierig⁹: Zu stark sind die Abweichungen der einzelnen Arbeitsschritte voneinander und zu unterschiedlich die jeweiligen Prioritäten und Schwerpunktsetzungen in den einzelnen Institutionen. Hinzu kommt außerdem noch, dass der Prozess der Modellierung und Ausgestaltung von Geschäftsprozessen selbst erhebliche Erkenntnisgewinne in der jeweiligen Organisation mit sich bringt, die für eine erfolgreiche Einfüh-

9 Ein leerreiches und transparent ablaufendes Beispiel für diese Bemühungen sind die Aktivitäten der AG E-Framework von DINI (Deutsche Initiative für Netzwerkinformation), siehe <http://www.dini.de/ag/e-framework/>, die sich zur Zeit darauf konzentriert, Verwaltungsabläufe in Hochschulen kooperativ zu modellieren.

rung neuer oder veränderter Geschäftsprozesse unverzichtbar sind. Kurz: eine einfache Übertragung „gegebener“ Modelle, die auf die individuelle Erarbeitung und Analyse verzichtet, dürfte im Regelfall nicht erfolgreich sein.

Die Informatik hat für die Modellierung und Notation von Geschäftsprozessen verschiedene methodische Herangehensweisen entwickelt, zum Beispiel die Ereignisgesteuerten Prozessketten (EPK), eine von Scheer und Mitarbeitern entwickelte Sprache zur Modellierung von Geschäftsprozessen¹⁰ und vor allem die Unified Modeling Language (UML) der Object Management Group (OMG), die in der Praxis heute dominierende (technische) „Sprache“ für die Modellierung von Daten, Verhalten, Interaktion und Aktivitäten.¹¹ Seit 2005 ebenfalls an die OMG angebunden ist die sich immer mehr verbreitende Business Process Modeling Language (BPML), eine XML-basierte plattformunabhängige Metasprache, die auch die graphische Umsetzung von Prozessen erlaubt.¹²

Legt man zum Beispiel UML als Syntax fest, sind noch methodische Festlegungen für die Herangehensweise zu treffen und es liegt nahe, sich für die vorbereitende Modellierung von technischen Abläufen in der Langzeitarchivierung am OAIS-Modell zu orientieren, das die prinzipiellen Aspekte im Umfeld der Langzeitarchivierung in funktionaler Perspektive beschreibt und an anderer Stelle dieser Enzyklopädie ausführlich dargelegt wird.¹³ Für den Bereich des Ingests einzubeziehen ist der Producer-Archive Interface Methodology Abstract Standard“ (CCSDS 651.0-B-1), der insbesondere Validierungsmechanismen und ihrer Einbindung in die Prozesskette betrachtet.¹⁴

Einzelne Funktionen lassen sich so vor der Folie bisher bereits gemachter Erfahrungen allgemein beschreiben. Beispiele für diese übergreifenden Basisprozesse sind (ich nenne nur Beispiele für unmittelbar aus dem Kontext der Langzeitarchivierung heraus relevante Prozesse):

- Plattform- und Systemübergreifendes Taskmanagement
- Daten- und Objekttransfer-Mimik (z.B. OAI, ORE)
- Extraktion und Generierung von Metadaten (METS, LMER)
- Validierung von Dokumentformaten (z.B. JHOVE)
- Persistente Adressierung und Zugriffsmanagement auf Objektebene
- Speicherprozesse

10 Keller (1992)

11 OMG Infrastructure (2007) und OMG Superstructure (2007)

12 Vgl. dazu <http://www.bpmi.org/>

13 Vgl. hierzu den entsprechenden Artikel von Nils Brübach / Manuela Queitsch / Hans Liegmann (†) in dieser Enzyklopädie als Kapitel 4: „Das Referenzmodell OAIS - Open Archival Information System“

14 Vgl. hierzu <http://public.ccsds.org/publications/archive/651x0b1.pdf>

- ID-Management
- Inhaltsauswahl / Basisrecherche
- Migrationsprozesse / Formatkonvertierungen
- On-the-fly-Generierung einer Bereitstellungsumgebung