

Archivierung von Unterlagen aus digitalen Systemen

24. Tagung des Arbeitskreises

„Archivierung von Unterlagen aus digitalen Systemen“ (AUdS)



Archivierung von Unterlagen aus digitalen Systemen

24. Tagung des Arbeitskreises

„Archivierung von Unterlagen aus digitalen Systemen“ (AUdS)

IMPRESSUM

Redaktion: Thomas Krämer

© Landschaftsverband Rheinland Köln, 2022

Alle Rechte vorbehalten

Satz und Layout: Jutta Weis, Pulheim

Umschlagfoto: Nationalarchiv Prag

URN: urn:nbn:de:hbz:5:2-1275621

Inhaltsverzeichnis

Vorwort	1
Rückblick nach vorne: Was alles nicht geklappt hat	3
<i>Christoph Schmidt, Landesarchiv Nordrhein-Westfalen</i>	
Collecting software of a government department: from strategy to practical challenges	7
<i>Anne Jolly, Service historique de la Défense</i>	
OAIS-Referenzmodell-konforme Software-Architektur	13
<i>Frank Obermeit, Landesarchiv Sachsen-Anhalt</i>	
Strategies in digital archiving at the Belgian State Archives. Between ideal and reality	27
<i>Rolande Depoortere and Johan Van der Eycken, Archives générales du Royaume, Bruxelles</i>	
Education and Training in Digital Recordkeeping in France	34
<i>Édouard Vasseur, Centre Jean Mabillon, École nationale des chartes – PSL, Paris, France</i>	
Die Ausbildung in (elektronischer) Schriftgutverwaltung und digitaler Langzeitarchivierung in Deutschland	42
<i>Sarah Rudolf, Gregor Patt, LVR-Archivberatungs- und Fortbildungszentrum, Pulheim</i>	
Die Entwicklung einer Datenbankstruktur für die Erschließung digitaler Archivalien im Stadtarchiv Nürnberg	53
<i>Annette Birkenholz, Stadtarchiv Nürnberg</i>	
Gewerbekatasterdaten als Pilotprojekt – Die digitale Langzeitarchivierung beim Kommunalen Rechenzentrum Niederrhein (KRZN) mit startext SORI	60
<i>Dr. Michael Habersack, Kreisarchiv Viersen, Fabian Näser, Startext GmbH</i>	
Archivierung von Daten aus Twitter: Das „Twitterarchiv“ über die Twitter-eigene Exportschnittstelle 2018–2021	67
<i>Stephanie Kortyla, Sächsisches Staatsarchiv</i>	
Qualitätssicherung bei wenig verbreiteten Datenformaten am Beispiel von Digitalen Topographischen Karten	74
<i>Oliver Brennecke, Antje Lengnik, Martin Vogel, Niedersächsisches Landesarchiv</i>	
Anonymisierung im Informationssystem des Historischen Archivs des ungarischen Staatssicherheitsdiensts. Umsetzung und Auswirkungen der COVID-19-Pandemie	82
<i>Zoltán Lux, Historisches Archiv der Staatssicherheitsdienste, Budapest</i>	
Vernetzen, Erschließen, Verwalten, Zugänglichmachen. Die Kulturerbedatenbank der Deutschsprachigen Gemeinschaft Belgiens.	86
<i>Vitus Sproten, Belgisches Staatsarchiv, Eupen</i>	
Durchsuch- & Extrahierbarkeit als signifikante Eigenschaften von PDF/A	94
<i>Claire Röthlisberger-Jourdan, Georg Büchler, KOST</i>	
Digitalisierung im öffentlichen Hochbau	97
<i>Flurina Camenisch, Staatsarchiv Graubünden, und Bernhard Stüssi, Staatsarchiv Zürich</i>	

Gemeinsamer automatisierter Ingest archivwürdiger Fachverfahrensdaten im Kommunalen Digitalen Langzeitarchiv Baden-Württemberg	104
<i>Miriam Eberlein, Stadtarchiv Heilbronn</i>	
Digitale Verbundarchive – Zu Chancen und Herausforderungen von archivischen Kooperationsmodellen in der digitalen LZA	109
<i>Thomas Krämer, LVR-Archivberatungs- und Fortbildungszentrum, Pulheim</i>	
XDomea-Schnittstelle TIZIAN: Datenmapping inkl. AIP-Bildung	116
<i>Daniel Baumann, Stadtarchiv München, Markus Schmalzl, Generaldirektion der Staatlichen Archive Bayerns</i>	
„Archifiltre“ – a tool of the French Social Ministries	122
<i>Anne Lambertt, Mission des Archives de France, Chloé Moser, Direction du numérique des Ministères Sociaux</i>	
Echtheitsnachweis to go.Werkzeug zur Integritätswahrung elektronischer Reproduktionen	126
<i>Tony Grochow, Konrad Meckel, Landesarchiv Thüringen</i>	

Vorwort

Am 22. und 23. März 2021 fand die 24. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ (AUdS) statt, die durch das LVR-Archivberatungs- und Fortbildungszentrum ausgerichtet wurde. Die Veranstaltung war in verschiedener Hinsicht besonders. Sie wurde ursprünglich für den 24. und 25. März 2020 geplant, musste aber wenige Tage zuvor aufgrund der sich rasant ausbreitenden Corona-Pandemie verschoben werden. Damit ging eine beinahe vollständige Reorganisation einher, da sich im weiteren Verlauf des Jahres 2020 zeigte, dass die Durchführung einer Präsenzveranstaltung auch im folgenden Jahr unwahrscheinlich erschien. Die Organisator/innen konzipierten daraufhin eine rein virtuelle Tagung, was dankenswerterweise die Zustimmung sowohl des Lenkungskreises, als auch der Moderator/innen und Vortragenden fand. Für alle Beteiligten bedeutete dies ein kleines Abenteuer, das mit damals neuen Herausforderungen verbunden war.

Schlussendlich konnte die Konferenz mit geringen inhaltlichen, aber größeren organisatorischen Anpassungen nachgeholt werden. Sie war die erste virtuelle Tagung des Arbeitskreises, aber sicherlich nicht die letzte. Mit über 240 Personen erreicht die Zahl der Teilnehmenden einen neuen Höchstwert. Die Durchführung einer digitalen Veranstaltung ermöglichte es, mehr Personen zuzulassen und erleichterte die Teilnahme. Gleichwohl veränderte sich dadurch der Charakter der Veranstaltung, die 1997 als Arbeitskreis mit einer überschaubaren Zahl von Teilnehmenden begann und sich immer mehr zu einer Konferenz im klassischen Sinn entwickelt hat.

Insgesamt decken die Tagungsbeiträge ein weites Spektrum an Thematiken von der Ausbildung, Übernahmeprojekten bis hin zu Dateiformaten ab. Die Referierenden stammen aus sechs europäischen Ländern mit sehr unterschiedlichen archivischen Landschaften und Traditionen. Auch die Herangehensweisen der jeweiligen Einrichtungen an die digitale Langzeitarchivierung differieren z. T. ebenso stark wie die Stellung, Spartenzugehörigkeit, Größe, Binnendifferenzierung und Erfahrung der Einrichtungen. Beides entspricht den Grundgedanken des Organisationsteams. Zum einen sollten auch mittlere und kleinere Einrichtungen die Möglichkeit erhalten, ihre Ideen, Vorhaben, Probleme und Ergebnisse zu präsentieren. Dies spiegelt die Überzeugung wider, dass die digitale Langzeitarchivierung nicht gelingen kann, wenn sie ein Projekt einiger weniger Institutionen bleibt, die über eine umfangreiche finanzielle und personelle Ausstattung verfügen. Dieser Ansatz fügt sich zugleich in die Tradition der AUdS-Tagungen ein, die von Beginn an Archivsparten-übergreifend konzipiert waren. Zum anderen sollten auch Erfahrungen und Entwicklungen der westeuropäischen Nachbarn in den Blick genommen werden, die bislang noch nicht im Fokus der Tagung standen.

Einen wichtigen Impuls setzte gleich zu Beginn Dr. Christoph Schmidt (damals Landesarchiv Nordrhein-Westfalen), der in seiner Keynote eindrücklich vor einem Scheitern durch Perfektionismus warnte. Den zweiten Tag eröffneten Claire Röthlisberger-Jourdan (KOST), Prof. Dr. Christian Keitel (Landesarchiv Baden-Württemberg) und Dr. Peter Worm (Stadtarchiv Münster). In diesem für AUdS neuen Format im Stil einer englischen Universitätsdebatte sollte eine durchaus provokante These anregend diskutiert und das Auditorium in die anschließende Aussprache eingebunden werden. Für alle Beteiligten war es ein Experiment, gerade, weil es sich um eine rein digitale Veranstaltung handelte. Die Reaktionen des Auditoriums und die rege Beteiligung erlauben jedoch ein positives Fazit. Frank Obermeit (Landesarchiv Sachsen-Anhalt) griff vielleicht unbewusst den Appell von Christoph Schmidt auf. Er wies darauf hin, dass die technischen Systeme und Organisationen, die sich derzeit der digitalen Langzeitarchivierung annehmen, auf Veränderung und Anpassung ausgelegt sein müssen („Build to change“). Kein derzeit verfügbares System kann von sich behaupten, ausentwickelt zu sein. Dies wurde auch durch die Beiträge der westeuropäischen Referent/innen deutlich. Beispielsweise warben Anne Lambert und Chloé Moser ausdrücklich dafür, sich an der Weiterentwicklung des Tools Archifiltre zu beteiligen.

Sich beide Notwendigkeiten, das Finden pragmatischer und umsetzbarer Lösungen sowie die damit verbundenen Verpflichtungen zur Weiterentwicklung, bewusst zu machen, mag dazu beitragen, die Herausforderung, die die Archivierung digitaler Informationen mit sich bringt, leichter anzugehen bzw. weiter engagiert zu verfolgen. Fast alle Beiträge sind in diesem Tagungsband versammelt. Die Präsentationen zu den Vorträgen sind auf der Homepage des Staatsarchivs St. Gallen abrufbar.

Die Organisator/innen haben für das Gelingen der Tagung zahlreichen Personen zu danken. Die Agentur Streamvent stellte, neben der technischen Plattform mit verschiedenen Tools zum Streamen der Beiträge und Austausch- bzw. Kollaborationsmöglichkeiten, vor allem viel Knowhow über die Gestaltung rein virtueller Konferenzen zur Verfügung. Das Bundessprachenamt übernahm die Simultanübersetzung der englischsprachigen Beiträge, hier seien insbesondere Frau Holl und Frau Marquardt für die sehr gute Übersetzungsleistung von nicht alltäglichen und bisweilen sperrigen Thematiken erwähnt. Die Technik für die Simultanübersetzung stellte die Firma Brähler zur Verfügung. Vor allem sei jedoch den Moderator/innen und Referierenden für die spannenden und interessanten Beiträge sowie den zahlreichen Teilnehmenden gedankt, die mit ihren Beiträgen zum Gelingen der Tagung beige-

tragen haben. Zuletzt hat das Organisationsteam allen Beteiligten des LVR-AFZ selbst herzlich zu danken, die in unterschiedlichen Funktionen einen Anteil an der Planung, Vorbereitung und Durchführung der Tagung(en) hatten.

Im Namen des Organisationsteams
Dr. Thomas Krämer

Monika Marner M. A.

Rückblick nach vorne: Was alles nicht geklappt hat.

Christoph Schmidt, Landesarchiv Nordrhein-Westfalen

1 Kurze Vorbemerkung zum Thema „Wahrheit und Wahrhaftigkeit“

Zu den ehernen Legenden des AUDES zählt folgende: Früher war alles besser. Oder zumindest: gemütlicher, familiärer und produktiver. Es gab noch nicht so viele Teilnehmer, und man sprach, wie in jeder guten Familie, ganz offen und einfühlend über alle Dinge, die einem Probleme bereiteten. So zumindest wurde mir die Geschichte des Arbeitskreises ganz vertraulich, in der Pause, unter vier Augen bei meiner allerersten AUDES-Tagung 2011 erzählt, und so habe ich im Jahr darauf angefangen, sie weiter zu erzählen.

Ich habe keine Ahnung, ob die Geschichte stimmt. Es wäre aber schön, wenn sie wahr wäre. Daher wuchs in den letzten Jahren an verschiedenen Stellen und in verschiedenen Köpfen der Wunsch, einmal einen kleinen Kontrapunkt zu all den jährlichen Siegesmeldungen zu setzen und ausschließlich über das zu sprechen, was alles in einem normalen deutschen Landesarchiv in puncto Langzeitarchivierung bislang nicht geklappt hat. Das Programmkomitee zeigte sich erfreut und gestattete einen Beitrag an prominenter Stelle.

Vorab möchte ich noch anmerken, dass es mit der bedingungslosen Wahrhaftigkeit natürlich auch in diesem Beitrag so eine Sache ist. Denn: Obwohl ehrlich im Detail, werde ich natürlich längst nicht über *alles* berichten, was schiefgegangen ist. Ausgelassen wurden (sehr wenige) echte Peinlichkeiten sowie (recht viele) Lappalien. Geblieben ist stattdessen alles, was (hoffentlich) zugleich unterhalten und belehren kann, ohne gleich als Wissenschaft gelten zu müssen. Ich beginne chronologisch.

2 Die Ungnade der verspäteten Geburt

Das erste, was im Landesarchiv NRW nicht geklappt hat, ist der zeitige Anfang. Der früheste mir bekannte Problem-aufriß zum Thema „Digitales Archiv“ in NRW stammt von Horst Romeyk, einem damaligen Kollegen im Düsseldorfer Hauptstaatsarchiv. Er veröffentlichte 1975 eine kleine Publikation mit dem Titel „EDV und Archive“¹, in dem er auch dem archivischen Umgang mit Verwaltungsdaten ein kleines Kapitel widmete. Die erste Echtübernahme in das digitale Archiv des Landes erfolgte 2012. Im Hinblick auf praktische Ergebnisse waren die dazwischenliegenden 37 Jahre zweifelsohne keine Erfolgsgeschichte.

3 Projekte, Projekte, Projekte

Ernsthafte Versuche, ein digitales Archiv aufzubauen, wurden beim Landesarchiv NRW tatsächlich erst ab 2009 unternommen. Vorher galt das im Archivwesen weitverbreitete Motto: „Das ist eine komplizierte Sache, und entweder, wir machen das richtig oder gar nicht.“ Und da wir uns an „richtig“ nicht herantrauten, blieb es erst einmal bei „gar nicht“.

Als das dann nicht mehr ging, wurde erst einmal ein Projekt ins Leben gerufen. Dieses trug den originellen Titel „DArch“ (für: *Digitales Archiv*), und der Projektplan hatte 42 Seiten. Seine Umsetzung war auf zunächst vier Jahre angesetzt, von denen ein guter Teil sich mit der Erarbeitung weiterer Anforderungen befasst hätte. DArch ist aus verschiedenen Gründen nie umgesetzt worden, und aus der heutigen Perspektive muss man sagen: glücklicherweise.

2010 dynamisierte sich die Entwicklung. Im Windschatten der Bankenkrise standen für kurze Zeit Geldmittel zur Verfügung, die relativ frei verausgabte und für den Aufbau des digitalen Archivs genutzt werden konnten. Da in dieser Situation wenig Zeit für ausgefeilte Projektpläne blieb, wurde in aller Eile eine Marktsichtung der vorhandenen Systeme vorgenommen. Dabei haben einige Alternativen nicht geklappt, unter anderem der Anschluss an DIMAG. Das Landesarchiv Baden-Württemberg, das DIMAG damals alleine entwickelte und betrieb, stand zu dieser Zeit in Verhandlungen mit Hessen und konnte uns nicht sagen, ob, unter welchen Bedingungen und wann eine Zusammenarbeit möglich wäre. Da bei uns aber Geld vorhanden, aber keine Zeit zu verlieren war, erhielt eine andere Lösung den Zuschlag, nämlich die Digital Preservation Solution der Firmen SER und HP (heute DXC).

Diese Lösung, die auf den damals bereits bestehenden digitalen Archiven beim Stadtarchiv Stuttgart und beim Bundesarchiv basierte, wurde in den Folgejahren aufgebaut, ausgebaut und betrieben. Dabei haben viele Sachen gut geklappt, einige Sachen dagegen nicht. Von diesen nun die vielleicht fachlich interessantesten Fälle.

1 Horst Romeyk, EDV und Archive. Ein Ratgeber, Düsseldorf 1975. Die zweite Auflage des kleinen Traktats von 1981 kann weiterhin für kleines Geld beim Landesarchiv NRW käuflich erworben werden: https://www.archive.nrw.de/sites/default/files/media/files/Veroeffentlichungsliste%20%282021-07_21%29.pdf (S. 20; abgerufen am 14.07.2023).

4 Der lange Schatten der Papierwelt

Da wäre zunächst das Thema „Datenmodell“. Generell ist das DIPS zugrundeliegende Datenmodell gelungen. Es ist fachlich durchdacht und so präzise formuliert, dass es auch technisch sauber umsetzbar ist. Kritischer zu betrachten ist in diesem Kontext allerdings die Definition der Archivpakete. In der Anfangsphase von DIPS wurde bewusst entschieden, den Begriff des „Pakets“ aus dem OAIS-Modell sehr wörtlich zu nehmen. Alle zu einem AIP gehörigen Dateien einschließlich Metadaten werden in DIPS in einen unkomprimierten Container-File gesteckt und als eine einzelne Paketdatei abgelegt. Wir erhofften uns damals von dieser Regelung eine größere Robustheit des Archivs gegenüber technischen Ausfällen. Tatsächlich sind die DIPS-Archivpakete vollständig selbsttragend und gegen einen Verlust durch Datenbankdefekte (zumindest theoretisch) weitgehend geschützt. Selbst bei einem weitreichenden Zusammenbruch des Datamanagements könnte man über das Filesystem alle verlorenen Informationen rekonstruieren.

Aus heutiger Sicht ist diese Konstruktion dennoch nicht empfehlenswert, und zwar aus drei Gründen: Erstens nämlich wurde hier bewusst oder unbewusst versucht, einem Problem der Bitstream-Preservation auf der Ebene des Datenmodells zu begegnen. Dem liegt ein diffuses Misstrauen gegenüber den eigenen Backup-Fähigkeiten und der hausinternen IT-Infrastruktur zugrunde, das besser auf andere Weise auszuräumen gewesen wäre. Zweitens wurde im Vorfeld keine adäquate Kosten-Nutzen-Abschätzung der Etablierung physischer Archivpakete durchgeführt. Physische Archivpakete machen fast alle Teilprozesse der digitalen Archivierung komplizierter, was sich insbesondere im Ingest wirklich großer Pakete, bei der Erzeugung neuer Repräsentationen und beim Update von Metainformationen auswirkt. Hier würde man heute sicherlich zu einer anderen Gewichtung kommen. Und drittens schließlich ist die in DIPS realisierte Vorstellung eines echt physischen Archivpakets doch arg analog und opfert die Flexibilität eines klaren Datenbanksystems dem eher ästhetischen Vorteil besserer Anschaulichkeit. Das ist vermutlich auch das dahinterliegende Grundproblem: Die Erschließung des vollen Potenzials einer neuen Technik wird oftmals dadurch verhindert, dass man zu Beginn dazu tendiert, bekannte, „altbewährte“ Strukturen einfach auf ein neues Medium zu übertragen. Das hat in diesem Fall zwar nicht zu einem echten Defekt geführt, macht aber einige Dinge langfristig komplizierter als notwendig.

5 Eigenartige Eigenschaften

Nicht geklappt hat es im Landesarchiv NRW bislang auch mit den signifikanten Eigenschaften. Und zwar nicht, weil der bestandserhalterische Ansatz falsch wäre, für den langfristigen Erhalt digitaler Objekte alle Eigenschaften zu bestimmen, die über sämtliche Migrationsstufen hinweg bewahrt werden müssen. Im Gegenteil: Ich bin weiterhin davon überzeugt, dass wir das, was wir erhalten und in der Nutzung bereitstellen wollen, möglichst genau auswählen und beschreiben müssen, um dann alle geeigneten Maßnahmen zum Erhalt von Information und Performance ergreifen zu können. Leider sind im Landesarchiv NRW jedoch bislang alle Versuche gescheitert, das Konzept der signifikanten Eigenschaften technisch produktiv umzusetzen.

Dafür gibt es vornehmlich drei Gründe.

1. In den meisten der mir bekannten Überlegungen zur Theorie und Praxis der signifikanten Eigenschaften ist der Begriff der „Eigenschaft“ nicht klar genug gefasst. Insbesondere gilt dies für die Unterscheidung zwischen inhaltlichen und funktionalen Eigenschaften eines Archivals, den technischen Eigenschaften einer Repräsentation sowie ihrem Verhältnis zueinander, und zwar insbesondere bei komplex strukturierten Objekten. Dies verhindert oder erschwert die praktische Erstellung von Eigenschaftskatalogen und Erhaltungsgruppen. Und ohne eine solche strukturell saubere und verlässliche Grundlage lässt sich eine technische Umsetzung nicht realisieren.
2. Um die signifikanten Eigenschaften eines digitalen Archivobjekts benennen zu können, muss man zunächst überhaupt seine Eigenschaften kennen. Dies mag in manchen Fällen einfach sein oder zumindest einfach erscheinen. Dummerweise ist aber eine ökonomisch umsetzbare, technisch automatisierbare Eigenschaftsanalyse keineswegs einfach zu etablieren. Ein Beispiel: Einer PDF-Datei ist ohne Weiteres nicht anzusehen, ob sie z. B. Text, Bilder, Text mit Bildern gemischt oder Text in Bildform enthält. Oder ob und wie ein Text formatiert ist, ob farbliche Hervorhebungen vorhanden sind, ob der Text durchsuchbar ist usw. Das Wissen darüber ist aber unabdingbar, um fachgerechte Bestandserhaltungsmaßnahmen planen und umsetzen zu können. An einem Konzept, wie man eine solche Analyse technisch zuverlässig und mit vertretbarem Aufwand durchführen könnte, sind wir bislang gescheitert.
3. Ein drittes gravierendes Problem betrifft die Definition von Signifikanz. Diese sollte nach gängiger Lehrmeinung entweder an den Interessen der Zielgruppen eines Archivs ausgerichtet sein oder sich am ursprünglichen Verwendungszweck in der Verwaltung orientieren. Beide Vorschläge sind allerdings unbefriedigend. Denn: Archive verfolgen in der Regel ganz andere institutionelle Ziele als ihre Datenproduzenten. So ist z. B. die Eigenschaft „kann weiter verarbeitet werden“ für eine lebende E-Akte von größter Bedeutung, im Archiv jedoch nur in Ausnahmefällen. Die Orientierung dagegen an *allen* möglichen Nutzerinteressen der Zukunft führt schnell in den Bereich unkalkulier-

barer Fantasie, zu Allem und Nichts. Und eine (ebenfalls denkbare) stärker fokussierte Zielgruppenorientierung gibt es innerhalb des Landesarchivs NRW bislang nicht. Sie zu definieren ist eine strategische Aufgabe, die über die Kompetenzen des digitalen Archivs hinausreicht.

Im Ergebnis blieb nach einigen Monaten ernsthafter Bemühungen die Erkenntnis, dass wir die genannten Probleme bis auf Weiteres nicht lösen konnten. Wir haben es daher erst einmal aufgegeben – mit dem unangenehmen Beigeschmack, eine gewichtige Hypothek auf die Zukunft aufgenommen zu haben.

6 Die Entdeckung der Langsamkeit

Der vermutlich deprimierendste, aber auch lehrreichste Fall von Nicht-Klappen betrifft die Aussonderung elektronischer Akten. Das Land NRW hat bereits frühzeitig Pilotversuche zur E-Aktenführung unternommen und dazu das Produkt „DOMEA“ der Firma Opentext eingesetzt. Dass dieses System von Haus aus keine funktionsfähige Aussonderungsschnittstelle hatte, war allgemein bekannt und ein Problem, das seiner Zeit alle Verwaltungen betraf, die DOMEA einsetzten.

Die mitbetroffenen staatlichen Archive erkannten die Problemlage relativ frühzeitig und meldeten ihre Entwicklungsbedarfe länderübergreifend an. Unter dem Dach der Produktpflege der Fa. Opentext wurde daraufhin 2009 eine Arbeitsgruppe eingerichtet, die die fachlichen Anforderungen für eine Aussonderungsschnittstelle erarbeitete. 2011 war das Aussonderungskonzept fertig, und 2012 legte die Fa. Opentext ein Umsetzungskonzept und einen Kostenplan vor. Von den ursprünglich an der Entwicklung beteiligten Archiven waren drei Landesarchive zu einer gemeinsamen Finanzierung bereit: Hessen, Mecklenburg-Vorpommern und NRW. Aus haushaltstechnischen Gründen verzögerte sich dann jedoch die Beauftragung der Fa. Opentext um ganze zwei Jahre. Als 2014 dann theoretisch alle Haushaltsmittel vorhanden waren, beschloss Mecklenburg-Vorpommern, aus dem Entwicklungsverbund wieder auszuschneiden. Hessen und Nordrhein-Westfalen nahmen daher Nachverhandlungen mit Opentext auf, um eine erträgliche Finanzierungslösung zu finden. Diese scheiterten an der mangelnden Kompromissbereitschaft des Herstellers, der inzwischen verständlicherweise kein gesteigertes Interesse daran hatte, in das inzwischen zum Auslaufprodukt gewordenen DOMEA noch viel Zeit und Geld zu investieren. Die Länder Hessen und NRW entschlossen sich daher, ihre jeweiligen IT-Dienstleister mit der Programmierung eigener Aussonderungsschnittstellen auf der Grundlage des ursprünglichen Aussonderungskonzepts zu beauftragen. Eine gemeinsame Entwicklung der beiden Länder wurde als nicht effizient genug abgelehnt. In NRW machte sich IT.NRW daraufhin an die Entwicklung. Dieses Vorhaben wiederum hatte jedoch auch mit gravierenden Schwierigkeiten zu kämpfen. Zum einen zeigte sich schon früh, dass die technischen Rahmenbedingungen für den Einsatz einer einheitlichen Schnittstelle in vielen Fällen an der Inhomogenität der unterschiedlichen Systeminstanzen scheitern würde. Denn bei der Einführung von DOMEA hatte man seinerzeit den einzelnen Behörden große Gestaltungsspielräume z. B. hinsichtlich der Objektstrukturen und der erlaubten Dateiformate gelassen. Eine fachgerechte Aussonderung wäre unter diesen Voraussetzungen nur mit gravierenden Vorarbeiten und Umstellungen an den jeweiligen Produktivsystemen verbunden gewesen. Zum anderen wurde die Entwicklung der Schnittstelle bei IT.NRW nicht prioritär verfolgt. Die anstehende Umsetzung des E-Government-Gesetzes und die Einführung eines neuen, einheitlichen Landes-DMS-Systems führten zu Personalengpässen und Verzögerungen.

Das hatte zur Folge, dass sich die Arbeiten an der Schnittstelle über mehrere Jahre hinzogen. Inzwischen ist die Schnittstelle zwar fertig, es gibt aber keine Behörden mehr, die aussondern wollen. Alle infrage kommenden DOMEA-Instanzen haben sich dazu entschlossen, ihre Altdatenbestände in das neue Landes-DMS zu migrieren und dann die Aussonderungsschnittstelle zu nutzen, die dieses System von Haus aus mitbringt.

Was lässt sich nun aus einem solchen Debakel lernen?

1. Die Archive müssen bei der Formulierung fachlicher Anforderungen und der Implementierung von Aussonderungsschnittstellen deutlich schneller werden, um überhaupt mit den Entwicklungszyklen der Produktivsysteme mithalten zu können.
2. Um schneller werden zu können, müssen die Archive in länderübergreifenden Projekten eine verlässliche, weniger auf Perfektion als vielmehr auf Effizienz und Flexibilität ausgerichtete Kultur des Miteinanders etablieren. Das heißt auch, dass man von Beginn an die Möglichkeiten gemeinsamer Finanzierung realistisch abschätzen sollte.
3. Es ist unabdingbar, archivische Belange bereits bei der Einführung archivrelevanter IT-Systeme möglichst massiv einzubringen. Das DOMEA-Debakel resultiert zum größten Teil aus einer ineffektiven Begleitung der DMS-Einführungen, weit vor Beginn der Arbeiten an der eigentlichen Aussonderungsschnittstelle.

7 Komplexe Komplikationen

Abschließend nun noch ein kurzer Blick auf das weite Feld des alltäglichen Übernahmebetriebs. Auch hier hat erstaunlich viel gut funktioniert. Von dem, was nicht, oder besser gesagt: nur so lala geklappt hat, seien hier noch zwei Bereiche angesprochen: die Archivierung kompletter Datenbanken und die automatisierte Übernahme von Webseiten. In beiden Fällen haben wir für unsere Übernahmeprojekte auf etablierte Standards gesetzt, nämlich auf das SIARD-Format für Datenbank-Dumps und auf WARC-Files für Webinhalte. In beiden Fällen haben sich trotz Einsatz verschiedener Softwaretools und einer Menge Arbeit nachhaltige Probleme nicht ganz bewältigen lassen. Der Hauptgrund dafür ist, dass die technische Komplexität und Vielfalt sowohl bei den Datenbanken als auch bei Webseiten inzwischen so groß ist, dass mit den vorhandenen Werkzeugen und vor allem mit den zur Verfügung stehenden Ressourcen fast nie vollständig zufrieden stellende Ergebnisse zu erzielen waren. Insbesondere dynamische Inhalte von Webseiten oder komplexere Verknüpfungen zwischen verschiedenen Domains lassen sich nur sehr bedingt oder auch überhaupt nicht in WARC-Crawls abbilden. Und über die vielfältigen Möglichkeiten, in Datenbanken Sachen zu machen, die man nie im Leben nach SIARD und dann wieder zurück transformieren kann, ließe sich problemlos ein sehr viel längerer eigener Beitrag schreiben. Das Problem der hohen Komplexität des Ausgangsmaterials hatte zur Folge, dass die Erstellung sowohl von SIARD- als auch von WARC-Files stets ein sehr hohes Maß an manueller und vor allem einzelfallspezifischer Nacharbeit erforderte. Und auch der Aufwand für die Qualitätssicherung war unangemessen hoch.

Leider haben wir für diesen Problemkomplex bislang keine befriedigende Lösung gefunden. Möglicherweise fehlen uns schlichtweg die notwendigen Ressourcen und Kenntnisse. Möglicherweise aber sind wir hier auch (früher als erhofft) an eine Grenze der Archivierungsstrategie „Migration“ gestoßen – nämlich aus etwas Fragilem, Komplexen etwas Stabiles, Einfaches zu extrahieren. Vielleicht täten und tun die Archive gut daran, das bislang meistens eher vernachlässigte Thema „Exit-Strategie“ nicht nur in Bezug auf einzelne technische Systeme, sondern auch auf ihre bestandserhaltende Gesamtausrichtung mitzudenken und sich mögliche Alternativwege der digitalen Archive von morgen zumindest offen zu halten.

8 Epilog

Und was ist nun die Moral von der Geschichte? Lässt sich aus all den Pleiten, Pech und Pannen überhaupt etwas lernen? Ist es sinnvoll und hilfreich, zu versagen?

Nein und ja. Denn auf der einen Seite sollte man sich davor hüten, in der Manier eines Glückskekse das Scheitern zu einem evolutionären Grundprinzip zu verklären. Fehler muss man vermeiden, wo immer man kann. Und wenn man schon aus Fehlern lernen möchte, dann am besten aus den Fehlern anderer Leute.

Auf der anderen Seite verändert das Nicht-Klappen tatsächlich die eigene Arbeitsperspektive. Man ist gezwungen, festgefressene Vorstellungen abzustößeln und manchmal auch die eigenen Grenzen zu erkennen und zu akzeptieren. Beides trägt dazu bei, gleichzeitig selbstbewusster, aber auch demütiger an eine spannende und mühselige Aufgabe heranzugehen und die eigene Neugierde nicht einschlafen zu lassen. In diesem Sinne hat das Scheitern im Nachhinein oftmals auch eine versöhnliche Dimension.

Einem altersweisen Kollegen gegenüber bemerkte ich vor einiger Zeit, dass wir in dreißig Jahren vermutlich über das, was wir heute tun, nur lachen würden. Darauf bemerkte er trocken: „Ja, aber damit wir lachen können, müssen wir es erst mal machen.“

Collecting software of a government department: from strategy to practical challenges

Anne Jolly, *Service historique de la Défense*

The development of information technologies has profoundly changed information management practices and the way we access and share information. The nature and proportion of some contents have increased significantly (statistical data, personal data, etc.). Archiving information contained in datasets and software archiving often leads to the selection but also to the (re)organisation of the original content and structure that is no longer fixed.

These unprecedented changes have raised questions about some archival concepts and practices related to the records and archives collections.

The archival community and preservation institutions have been interested for several years in the organisational and technical impacts of the use of new technologies on records¹. Archivists and historians alike have also grasped the potential for research offered by the emergence of some types of digital content². On the other hand, reflections remain marginal or are rarely formalised with regard to two aspects of digital archives collections, and more precisely the acquisition of information system content:

- The relevance of traditional concepts and approaches to appraising and selecting digital data and documents³.
- The impact of the archivist's choices on the organisation and structure of the data he collects, and thus on the representation of the information and its reception.

I would like to share some reflections and feedback on the collection of Information Systems (IS) content at the level of a ministerial department. It is the phase prior to the transfer of digital material into a repository that interests us (the initial negotiations with entities, the definition of management and archival requirements...), apart from the technical issues of data export and format that are already the subject of numerous publications.

In this perspective, the strategy developed by the „Service historique de la Défense“, the institution in charge of preserving the archives of the French ministry of Defense, will be presented and reviewed. Finally, the need to develop some concepts and to reflect on the role of the archivist in the selection and organisation (or reorganisation) of the data collected will be discussed.

1 How does the increasing use of software impact on the work of archivists?

The use of information technologies in public institutions as well as companies has led to two major changes that impact on the archival methodology and practices.

The “balkanization” of the storage of digital resources

First, the digital materials created by entities are often managed in several software. What we can call „a balkanization“ of management and storage of digital resources. Indeed, despite efforts to centralize administrative processes in a few software applications, many information systems are still designed and used for one or more specific processes. This situation can also be explained by the history of technologies and the progressive equipment of organisations.

The replication of data

Then, archivists have to deal with a multiplication of data. More precisely, many digital resources are partially or fully replicated and reused in different software. Of course, this phenomenon existed before the use of digital technologies but it has intensified with the ease offered by digital tools.

1 This is one of the most dynamic areas of thinking in digital preservation organisations and institutions. One thinks of the numerous monitoring reports of the Digital Preservation Coalition (the DPC Technology Watch publications, online at: <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>) (retrieved on: 14th July 2023), the work carried out by the Swiss Federal Archives on the SIARD format or the preservation policy published by preservation institutions such as the Bibliothèque nationale de France, the Library of Congress or the Harvard Library.

2 In particular with regard to social networks: Aurore François, Anne Roekens, Véronique Fillieux, Caroline Derauw, eds, *Pérenniser l'éphémère. Archivage et médias sociaux*, Louvain-la-Neuve 2018, 235 p. The DPC Watch Report „Preserving transactional data“ highlights the potential of databases archiving for „futur developments in consumer analytics and in academic“. However, the report mainly focuses on preservation concerns. Sara Day Thomson, *Preserving transactional data*, DPC Technology Watch Report 16-02 May 2016, 44 p., online at: <https://www.dpconline.org/docs/technology-watch-reports/1525-twr16-02/file> (retrieved on: 14th July 2023).

3 For a reflection on long-term developments: Edouard Vasseur, *French archivists, the management of records and records management since the nineteenth century: are French recordkeeping tradition and practice incompatible with records management?*, in : *Archives and Manuscripts*, 49 (1-2) : 1–26, November 2020, online at: DOI: 10.1080/01576895.2020.1828108 (retrieved on: 14th July 2023).

Most of the time the replication occurred at the various stages in the lifecycle of the digital data. This reproduction is often due to the use of different software to provide functionalities related to the information lifecycle; such as creation of digital materials, collaborative creation, data sharing, validation, dissemination and medium-term preservation, as shown in the graph below.

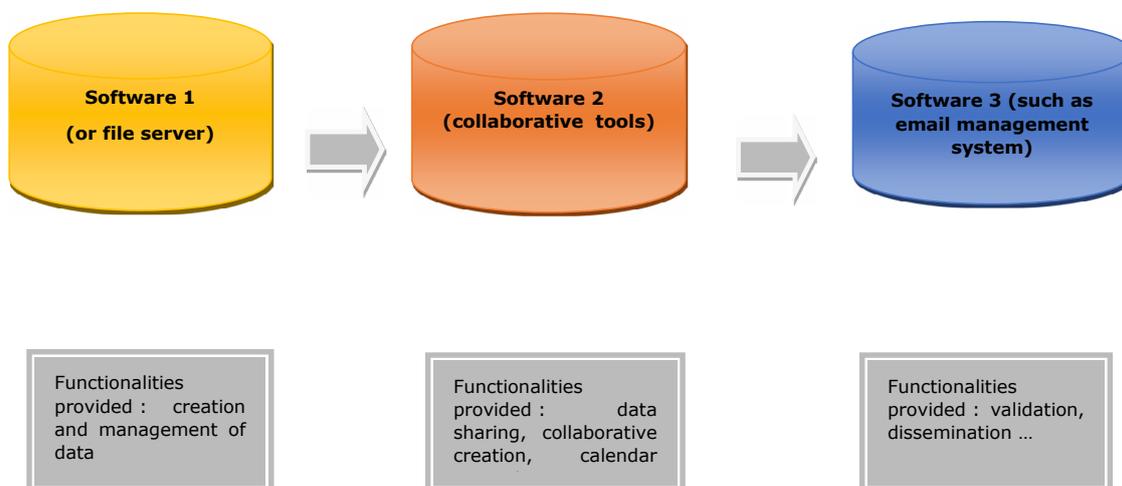


Fig. 1: Example of a dataflow during its lifecycle. Graph: Anne Jolly.

For several years now, processes have been simplified with the design of systems that provide more comprehensive workflows.

On the other hand, a new model of data management and use has emerged. Additional repositories and systems have been developed for the re-use of data, including statistical analysis and artificial intelligence tools. These repositories centralise data from different sources. Today, it can be seen that the transfer of content from several softwares to these data repositories is carried out without consulting the archivists. There is a risk of data loss (there is almost always a selection of transferred data and original content may be deleted) and a loss of context of data creation.

In addition to the traditional archival approach, it is therefore necessary to:

- Identify the various applications and storage systems used by an entity,
- Identify the interactions of the entity with the different software and digital tools available for the same functional area.

In this context, what strategy for identifying relevant software should be designed to meet these challenges? There is of course no single strategy. Indeed, the number and use of technologies can vary significantly from one entity to another. Several archivists working within French ministries have defined or implemented strategies. We would like to share the one designed by the “Service historique de la Défense”.

2 The unclassified software collection strategy at the “Service historique de la Défense”

In order to adapt to the increasing use of technologies, the “Service historique de la Défense” has adapted its procedures and methods in order to archive unclassified software 8 years ago. He defined a specific collection strategy.

The basis of the strategy

The „Service historique de la Défense“ set up a strategy focused on taking over unclassified software of permanent value. Its strategy is based upon the software managing policy of the ministry. Proper archival processes are integrated in the software lifecycle and specific procedures have been designed by the archivists.

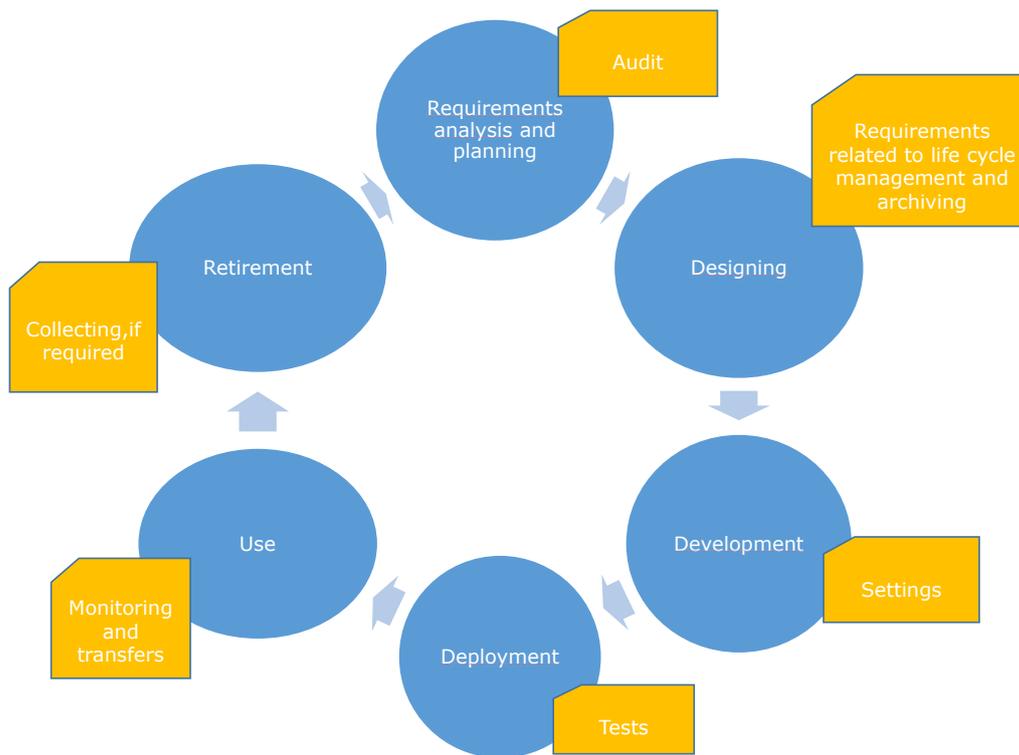


Fig. 2: The integration of archival procedures into software projects. Graph: Anne Jolly.

The scope of intervention includes all kind of software applications and management systems: from databases to electronic document management systems to email management systems and web sites. The strategy doesn't include office folders stored on a file server, if they have no link with the software content.

The implementation of the strategy is supported by a specific professional organisation. Some digital archivists are responsible for analysing defined software content. They are involved in the major phases of the software lifecycle: creation, evolution, replacement. They are responsible for the assessment, appraisal and selection of software content. The objectives are to identify and secure essential information and data of legal, administrative or historical value and to define the corresponding functional and technical requirements, which must be implemented in the software.

A methodology for assessing and prioritising the software

To deal with the many software projects and existing software, an audit grid has been produced. It allows the identification of high-value software content and enables to define priorities and the appropriate level of intervention.

The audit grid is based on the following criteria:

- **The context** (the producer(s), his/their position in the ministry, the mission concerned),
- **The process(es) involved.** The review is based on the following points and questions:
 - is the process fully performed by the software? Are any documents created and stored in the information system (IS)?
 - The data lifecycle (provenance and destination),
 - The type of needs for access and use of the data over time,
 - The need to ensure the integrity, authenticity and reliability of data (including technical constraints).
- **The content value (estimation).** We conduct an initial assessment of the administrative, legal and historical value of the software on the basis of documentation and exchange with the stakeholders. The overall retention period is also estimated.
- **The technical and organisational constraints** are identified. In particular, we pay attention to the monthly or annual volume managed, the storage capacity of the system, the type of development (external or not) and timing, the kind of software (open source or proprietary).

This analysis makes it possible to determine the issues associated with the software and its content and the type of intervention required. In practice, three main categories of IS can be distinguished, depending on the type of

associated data management and the level of intervention required. The following graph illustrates the characteristics of these categories.

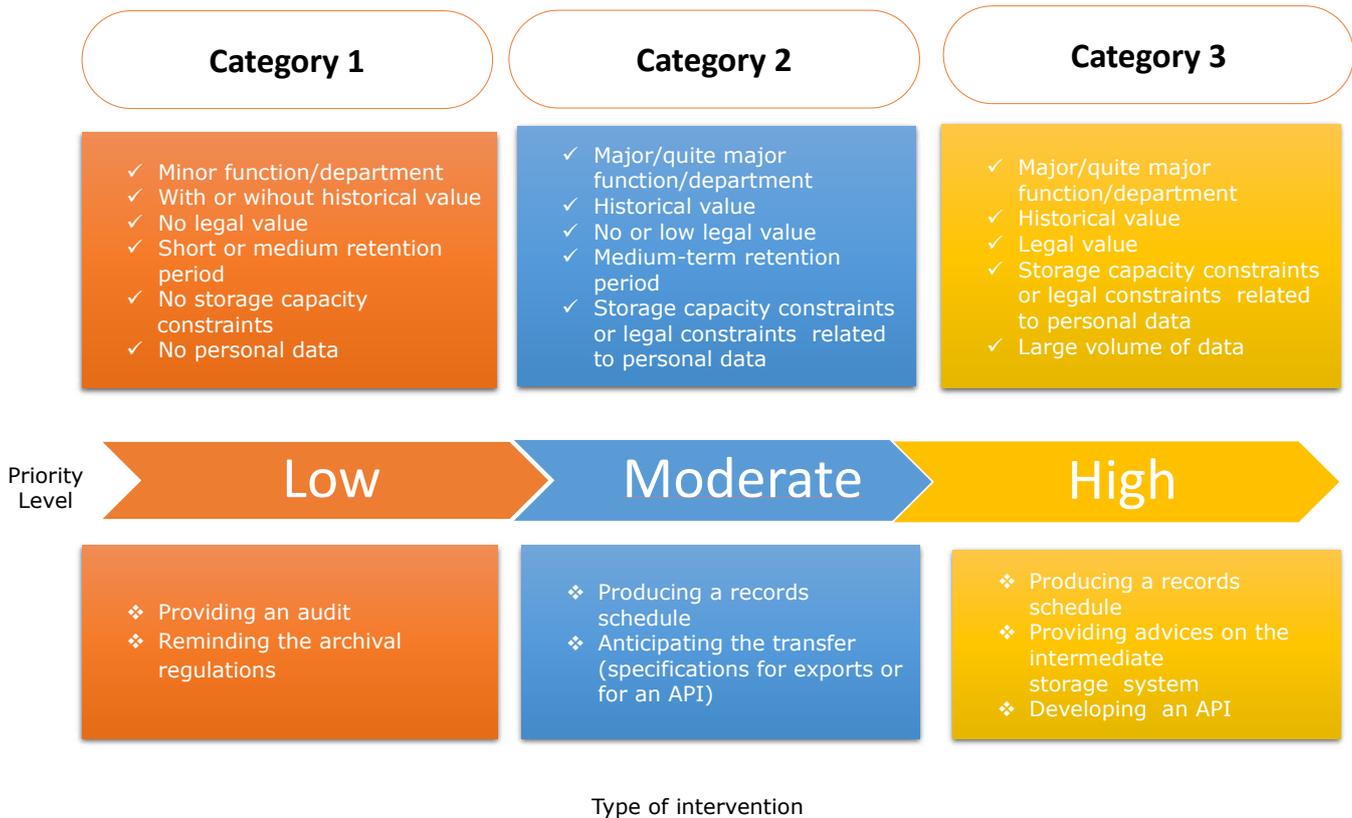


Fig. 3: Categorisation of IS project types and indication of the corresponding type of intervention. Graph: Anne Jolly.

As regards the first category of software project, the digital material can remain in the system safely for as long as users need it. If the data is of historical interest, it can be exported without technical difficulties a few months before the system is removed. In such cases, the intervention is limited to advice and a reminder of the regulations, and later, if necessary, the transfer of all or part of the data to the digital repository system.

With regard to the second group of IS projects, their content is partly of permanent value and has a low legal value. However, storage constraints or time limits on the retention of personal data requires an intermediate storage solution (or restricted access in the IS at the end of their useful life). The appraisal and the type of data transfer expected must be anticipated to avoid data loss. It is necessary to integrate functionalities into the software to enable compliant transfer into the preservation environment.

Concerning the third category, there is a double challenge of integrity and sustainability over time. It implies a significant involvement of the archivist, the entity(ies) concerned, the IS project manager and the electronic archiving system manager. The legal and historical value on the one hand, and the extended medium-term retention on the other hand, often require the development of an API for the transfer into the archive and, if necessary, the communication of archives.

3 Review of the strategy and its implementation

The strategy has led to positive results insofar as it has allowed the integration of archiving into a particular governance, that of the IS.

The strategy makes it possible to deal with some specific challenges of software archiving. Software archiving means above all being part of a project. New external factors place constraints on the working process of the archivist.

The first one is the project schedule. The software project deadlines are in addition to the usual deadlines and events that lead to the archivist's intervention, such as administrative reorganisations, the absence of regular archive transfers... This is why the capacity to evaluate and prioritise the projects is fundamental. Software archiving involves the responsibility and availability of archivists.

The financial constraints of the project are the second point. Archivists must assert that the costs of archiving are integrated into the IS project budget. For this purpose, they must define the appropriate needs for archiving and argue

for the costs to be taken into account from the start of the project.

The third challenge is technical. It is necessary to define the exports or interfacing requirements in the project phase in which the expression of needs is possible.

A strategy makes it possible to anticipate and deal with these issues. However, the way of intervention is often too dependent on the project schedule. In order to assess the interest of the content of an IS, an overall analysis of the production of one or more entities in the area concerned is necessary. To ensure that all the relevant records are identified, it is necessary to spend time on the definition of the records collection strategy and on a plan that integrates the IS as well as all paper and digital archives of the entities within its scope. It is necessary for this collection plan to include in its prioritisation criteria, for each existing functional area, both the traditional archiving criteria and the significance, the value and the priority of IS, whether they are future software projects or upcoming retirements.

Apart from these organisational issues, the challenge of collecting digital archives is much more than a global reflection that must involve other professional actors (archivists, historians, lawyers, computer scientists) at two levels :

- A reflection on concepts, even legal notions, related to archival holdings.
- Another reflection on the new role of the archivists in the choice of the structure of the information they preserve and the impact of this choice in the transmission and understanding of digital archives for current and future generations.

4 New ways of creating, managing and using digital data that require rethinking some archival concepts

The concept of the public records created by an entity is partially transposable to the IS context

The notion of records created or received by a public body as it exists in the archival tradition and the legal framework is partly inoperative for the apprehension of datasets managed in some IS.

For example, some systems are designed to allow multiple actors to enter, validate and transmit data to higher authorities. These authorities and some intermediate actors can modify or request modifications of these data. A single centralised system provides all the necessary workflows and there is only one copy of the data. Only the access varies according to the type of actor. In such cases, it is difficult to determine which entity is considered responsible for the data. Retention periods must be determined according to the needs of the different actors, which may vary.

Collaborative data creation and sharing tools also allow different entities or departments of an entity to collaborate on particular areas within a common system. Although a special entity managing the space is clearly designated, this organisational choice only imperfectly reveals the role of the different contributors and actors in the creation of data.

The difficulty of identifying the organisation or department responsible for the data in certain IS projects makes the archivist's work more complex. Which entity should be considered as his interlocutor? What are the responsibilities of each actor in terms of management and preservation of public data?

The second challenge of IS archiving is related to the constitution of the heritage that we wish to transmit, more precisely the ways of restitution of the collected information that we wish to retain

On the criteria for the structure of archive-relevant data

This question is not trivial. I am not dealing here with the choice of format made for preservation purposes, even though they may have an impact on decisions in terms of the form and restitution of collected data.

It is a question of taking into account the disappearance of the representation of information in a fixed form, as it has existed since the invention of writing (wax tablets, volumes, codex). Some researchers and records managers are carrying out research on this issue, as shown by a recent seminar organised in Paris, on 8 July 2020, by the Archives nationales [de France], the Jean Mabillon Centre (Ecole nationale des chartes) and the Dicen-IDF laboratoire (Cnam Paris) on the subject of archiving and using the datasets⁴.

Information systems centralise aggregates of more or less structured data. They display a selection of data on screens (man-machine interfaces) to show a set of information according to a logical structure and layout defined when the IS was designed based on user needs⁵. There are exceptions; some systems (such as data visualization systems, particularly for statistical and prospective uses) allow users to define the content and display what they wish to obtain.

Mostly, the systems allow users to view a set of information in a fixed form (electronic form, succession of fixed field

4 National Archives, the Jean Mabillon Centre (École nationale des chartes) and the Dicen-IDF laboratory (Cnam Paris), *Le tout et la partie: Lagrégation de données numériques et leur conservation pérenne*, seminar organised on 8 July 2020. The videos of the talks are available online at: <https://nparchive.hypotheses.org/tag/donnees-agregees> (retrieved on: 14th July 2023).

5 The records collection policy of the UK National Archives briefly recalls this evolution in the way existing information is accessed in the form of data in the IS (see 3.2 „Datasets“). National Archives of the United Kingdom, *Records collection policy*, Last reviewed October 2019, p. 8, online at: <https://www.nationalarchives.gov.uk/documents/information-management/records-collection-policy.pdf> (retrieved on: 14th July 2023).

displays...). The archivist may retain only what he considers relevant and will in fact define a digital archive that is different, in content and form, from those created and seen by the software users at the time of use of the system.

In such a technological context, the field of intervention of the archivist evolves. He no longer selects fixed documents according to the interest of their content. In addition to selecting valuable information, the archivist can define how to combine it, and how to aggregate and structure the data containing that information.

Archivists are already thinking in practice about the contextual elements to be archived on IS in order to document, when relevant, and trace the uses and ways of access to IS data by our contemporaries. The archiving of IS themselves remains little practised in French archives either because the IS are not of interest or because of technical difficulties.

It seems necessary to consider the impact of these choices made by archivists when collecting the datasets in terms of the transmission of the collected heritage to future generations. Understanding a given society obviously requires, in particular, knowledge of its relationship with the written content and representations that surround it. These considerations deserve more discussion between historians, archivists and records managers. These reflections should allow archivists to make informed decisions about digital records collection.

Bibliography

Digital preservation coalition, Technology Watch publications, online at: <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports> (retrieved on: 14th July 2023).

Aurore François, Anne Roekens, Véronique Fillieux, Caroline Derauw, eds, *Pérenniser l'éphémère. Archivage et médias sociaux*, Louvain-la-Neuve 2018, 235 p.

Sara Day Thomson, „Preserving transactional data“, DPC Technology Watch Report 16-02 May 2016, 44 p., online at: <https://www.dpconline.org/docs/technology-watch-reports/1525-twr16-02/file> (retrieved on: 14th July 2023).

Edouard Vasseur, „French archivists, the management of records and records management since the nineteenth century: are French recordkeeping tradition and practice incompatible with records management?“, in *Archives and Manuscripts*, 49(1-2) : 1-26, November 2020, online at: DOI: 10.1080/01576895.2020.1828108 (retrieved on: 14th July 2023).

OAIS-Referenzmodell-konforme Software-Architektur

Frank Obermeit, Landesarchiv Sachsen-Anhalt

Der Autor hat im Rahmen seiner Mitarbeit im DIMAG-Entwicklungsverbund, als der das DAN vertretende Informatiker und als Mitglied der AG „Stringente Gesamtarchitektur“ einen Software-Architekturvorschlag (zukünftig nur Architekturvorschlag) erarbeitet, mit dem das OAIS-Referenzmodell software-technisch vollständig umgesetzt werden kann. Dieser Architekturvorschlag wurde Anfang Januar 2020 dem DIMAG-Entwicklungsverbund übergeben und konnte wegen Corona erst 2021 dem Arbeitskreis AUdS¹ vorgestellt werden.

Die Grundsätze des Architekturvorschlages wurden bereits in Prag 2019² vorgestellt. Seit September 2020 ist im Landesarchiv Sachsen-Anhalt eine den Architekturvorschlag vollständig umsetzende und nachnutzbare Appliance verfügbar. Entwicklungszeiten verkürzten sich auf ein Zehntel.

Dieser Tagungsbandartikel richtet sich an Informatiker, Archivar/innen und Entscheider/innen und soll ein gemeinsames Verständnis der Gesamtarchitektur bei allen fachlich, technisch und organisatorisch Verantwortlichen vermitteln. Seinen Zweck erfüllt dieser Artikel, wenn die Anspruchsberechtigten (engl. Stakeholder) ihre Interessen erfüllt sehen. Der Artikel basiert vorrangig auf dem im Januar 2020 dem DIMAG-Entwicklungsverbund übergebenen Software-Architektur-Vorschlag³.

1 Software-Architektur und -Design

Eine **Software-Architektur** beschreibt die grundlegenden Komponenten und deren Zusammenspiel innerhalb eines Softwaresystems und bezieht sich dabei typischerweise auf die größeren Strukturen. Sie befasst sich mit der Art und Weise, wie mehrere Softwareprozesse zur Erfüllung ihrer Aufgaben zusammenarbeiten. Das **Software-Design** (Software-Entwurf) bezieht sich auf die kleineren Strukturen und befasst sich mit dem internen Design eines einzelnen Softwareprozesses. In der Software-Architektur werden nicht-funktionale Entscheidungen gefällt und von den funktionalen Anforderungen getrennt. Im Software-Design erfolgt die Beschreibung der funktionalen Anforderungen. Die Software-Architektur dient als Blaupause für ein System. Sie bietet eine Abstraktion, um die Systemkomplexität zu verwalten und einen Kommunikations- und Koordinationsmechanismus zwischen den Komponenten zu etablieren. Die Struktur des Systems wird aufgedeckt, doch die Details der Implementierung werden ausgeblendet. Das Software-Design bietet einen Designplan, der die Elemente eines Systems beschreibt, sowie die Art und Weise, wie sie zusammenpassen und zusammenarbeiten, um die Anforderungen des Systems zu erfüllen, und fungiert als Blaupause für den Entwicklungsprozess.

1.1 Einführung

Der Architekturvorschlag wird auf einem hohen Abstraktionsniveau und in natürlicher Sprache beschrieben. Das Dokument wurde in Anlehnung an das Hewlett-Packard-Dokument „A Template for Documenting Software and Firmware Architectures“⁴ strukturiert und das Template selbst nach dem 4+1 View Model von Krutchen⁵ aufgebaut. Auf die Benutzung der Beschreibungssprache UML wurde zum besseren Verständnis des Dokuments verzichtet. Der eigentliche Architekturvorschlag basiert auf den Erfahrungen und Erkenntnissen aus dem im Landesarchiv Sachsen-Anhalt⁶ durchgeführten Projekt zur Umsetzung einer auf BPMN und REST basierenden OAIS-Referenzmodell-konformen Software-Architektur⁷.

1 Online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2021/strategien-der-daten%C3%BCbernahme/Obermeit_BPMN_REST_Framework.pdf (abgerufen am: 14.07.2023).

2 Online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/camp-1-d/D_01_Obermeit_BPMN-REST-OAIS-Architektur-Modell-2019-03-13-014_skript.pdf (abgerufen am: 14.07.2023); https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/camp-1-d/D_01_Obermeit_BPMN-REST-OAIS-Architektur-Modell-2019-03-13-014.ppsx (abgerufen am: 14.07.2023).

3 Siehe auch den online verfügbaren Videovortrag auf der oben genannten Seite.

4 Michael A. Ogush, Derek Coleman, Dorothea Beringer, A Template for Documenting Software and Firmware Architectures, Version 1.3, 15-Mar-0, Online unter: https://www.cs.helsinki.fi/group/os3/HP_arch_template_vers13_withexamples.pdf (abgerufen am: 14.07.2023).

5 Online unter: https://en.wikipedia.org/wiki/4%2B1_architectural_view_model (abgerufen am: 14.07.2023).

6 Online unter: <https://landesarchiv.sachsen-anhalt.de> (abgerufen am: 14.07.2023).

7 Siehe Fn. 2.

1.2 Systemzweck-Beschreibung

Die Softwarearchitektur soll die Umsetzung des gesamten OAIS-Referenzmodells ermöglichen. Das auf der Software-Architektur beruhende System soll von Archiven beliebiger Größe genutzt werden können. Das System soll beliebig skalierbar sein und die Umsetzung aller Schutzbedarfskategorien ermöglichen.

1.2.1 Nicht-funktionale Anforderungen

Nicht-funktionale Anforderungen können in drei Arten unterteilt werden:

1. Service-Qualitäten, z. B. Leistung, Durchsatz, Benutzerfreundlichkeit, Sicherheit
2. Bedingungen und Einschränkungen, z. B. Plattformen, Bedingungen für die Qualität der Dienstleistung
3. Grundsätze, z. B. wenig Eigenentwicklung, Möglichkeit des Supports bei Open-Source- oder Community-Produkten

1.2.1 Service-Qualitäten

- Autorisierung und Authentifizierung im Front- und Backend-Bereich
- Single Sign-on
- Mandantenfähigkeit
- Verteilbarkeit
- Skalierbarkeit
- Abrechenbarkeit von Leistungen

1.2.2 Bedingungen und Einschränkungen

- Web-Client-Server-Architektur
- Allgemeingültige Standards und keine proprietären Vorgaben
- BPMN 2.0 als Standard zur Prozess-Modellierung und -Abarbeitung
- SIP, AIP und DIP müssen jederzeit aus sich heraus wieder herstellbar sein
- Backup und Recovery sind nicht Bestandteil des Architekturvorschlages, da eine Lösung von den Gegebenheiten des IT-Dienstleisters abhängt

1.2.3 Grundsätze

- Das System wird in wiederverwendbare und logische Komponenten mit klar definierten Kommunikationsschnittstellen geteilt.
- Das Rechte- und Rollenkonzept des Archivs muss umsetzbar sein.
- Alle Komponenten des Systems arbeiten serviceorientiert.
- Komponenten können nur über deren öffentliche Schnittstellen benutzt werden.
- Die Software-Komponenten basieren auf markterprobten Produkten, die als Community respektive Open-Source-Produkte verfügbar sind, jedoch die Möglichkeit des professionellen Supports bieten.
- Die Softwarearchitektur muss auf technologische Änderungen und auf Änderungen von Standards vorbereitet sein – *Build to Change Instead of Building to Last*⁸.
- Es werden Technologien mit offenen Standards zur Erstellung moderner Webanwendungen verwendet – *Use open-standards technologies to build modern web apps*⁹.

2 Strukturelle Beschreibung – Logical View

„Logical View“ beschäftigt sich mit der Funktionalität des Systems für den Endnutzer. In diesem Abschnitt wird die statische Architekturstruktur in Bezug auf ihre logischen Komponenten und deren Verbindungen beschrieben. Dabei sind logische Komponenten Verantwortungseinheiten auf einer bestimmten Abstraktionsebene. Komponenten auf einem höheren Abstraktionsniveau werden oft als Subsysteme bezeichnet. In diesem Dokument bezeichnet der Begriff Komponente immer eine logische Komponente.

8 Online unter: https://www.tutorialspoint.com/software_architecture_design/key_principles.htm (abgerufen am: 14.07.2023).

9 Online unter: <https://developer.ibm.com/technologies/web-development/> (abgerufen am: 14.07.2023).

2.1 Übersicht

Abstract: Der Architekturvorschlag beschreibt, wie eine Open Archival Information System (OAIS)¹⁰-konforme Software-Architektur mit Business Process Model and Notation (BPMN)¹¹ und Representational State Transfer (REST)¹² umgesetzt werden kann. Das auf den Standards OAuth2¹³ und Open ID Connect¹⁴ basierende Identity Access Management¹⁵ bildet den Schlussstein der ganzheitlichen Umsetzung des OAIS-Referenzmodells. Die Software-Architektur kann vorrangig mit Open Source Software, Community Software und Oracles¹⁶ no-cost-features unter Benutzung einer Container-Virtualisierungssoftware umgesetzt werden. Es wird ein für den individuellen Einsatz konzipiertes System vorgeschlagen, das out-of-the-box einsetzbar sein wird. Der Leitgedanke des vorliegenden Vorschlages lautet: Standards und etablierte Produkte einsetzen und Eigenentwicklungen auf den archivarisches Kernbereich beschränken.

Gesamtüberblick: Das OAIS-Referenzmodell beschreibt neben den Außenbeziehungen die sechs Functional Entities Ingest, Data Management, Archival Storage, Access, Preservation Planning und Administration. Der Archival Storage ist die zentrale Functional Entity; alle anderen Functional Entities arbeiten direkt oder indirekt mit dem Archival Storage. Alle Functional Entities stellen Dienste und Funktionen bereit, die durch Prozesse miteinander verknüpft werden. Diese Prozesse werden als Workflows bezeichnet. Der seit 2011 existierende Open Management Group¹⁷ (OMG) Standard BPMN 2.0 spezifiziert die grafische Modellierung von Workflows. Letztendlich haben alle Funktionen und Prozesse die Aufgabe, die Submission Information Packages (SIP) vom Producer entgegenzunehmen, Archival Information Packages (AIP) für die dauerhafte Aufbewahrung zu bilden und als Dissemination Information Packages (DIP) dem Consumer bereitzustellen. Dauerhaft werden nur die AIP im Archival Storage aufbewahrt und verwaltet. Wird davon ausgegangen, dass SIP und DIP in Storages aufbewahrt werden, wenn auch nicht dauerhaft, dann sind dafür ebenfalls Archival Storages notwendig und sollten wie die Functional Entity Archival Storage betrachtet und konzipiert werden. Mit REST-Webservices können Dienste und Funktionen zur Benutzung der Archival Storage gekapselt und mittels BPMN miteinander verknüpft werden.

BPMN verwendet das Konzept transaktionaler Subprozesse in Kombination mit Kompensierungsereignissen und -aktivitäten zur Transaktionssteuerung in verteilten Systemen¹⁸.

Graphical User Interfaces und Rechteverwaltungen unterliegen häufigen Veränderungen. Dafür wird es nie dauerhafte Lösungen geben, also müssen diese als austauschbare Komponenten in einer Software-Architektur konzipiert werden. Funktionen, Dienste und Prozessabläufe unterliegen ebenfalls regelmäßigen Änderungen. Diesen Aspekten muss eine OAIS-konforme Software-Architektur gerecht werden. Die Dokumentation und Nachvollziehbarkeit abgearbeiteter Workflows muss außerdem berücksichtigt werden. Der Autor ist davon überzeugt, dass sich die Container-Virtualisierung, wie die Betriebssystem-Virtualisierung vor zwanzig Jahren, auch bei den IT-Dienstleistern der Archive durchsetzen wird, denn durch ökonomische Zwänge wird eine bessere Ausnutzung der vorhandenen Hardware-Ressourcen unumgänglich sein. Besonders hervorzuheben sind die Vorteile der Containerisierung für die Softwareentwicklung und -bereitstellung. Der vorliegende Architekturvorschlag ist allgemeingültig und bietet alle notwendigen Möglichkeiten für den individuellen Einsatz.

10 Online unter: <https://public.ccsds.org/pubs/650x0m2.pdf> (abgerufen am: 14.07.2023).

11 Online unter: <http://www.bpmn.org> (abgerufen am: 14.07.2023).

12 Online unter: https://en.wikipedia.org/wiki/Representational_state_transfer (abgerufen am: 14.07.2023).

13 Online unter: <https://oauth.net/2/> (abgerufen am: 14.07.2023).

14 Online unter: <https://openid.net/connect/> (abgerufen am: 14.07.2023).

15 Online unter: https://de.wikipedia.org/wiki/Identit%C3%A4tsmanagement#Identit.C3.A4tsmanagement_von_Unternehmen (abgerufen am: 14.07.2023).

16 Online unter: <https://www.oracle.com/de/index.html> (abgerufen am: 14.07.2023).

17 Online unter: <https://www.omg.org/> (abgerufen am: 14.07.2023).

18 Online unter: <https://blog.holisticon.de/2011/05/transaktionen-bpmsoa/> (abgerufen am: 14.07.2023).

2.2 Beschreibung der Komponenten

Der Autor hat sich für ein komponenten- und schichtenbasiertes Modell entschieden, welches nachfolgend vom OAIS-Referenzmodell ausgehend entwickelt wird.

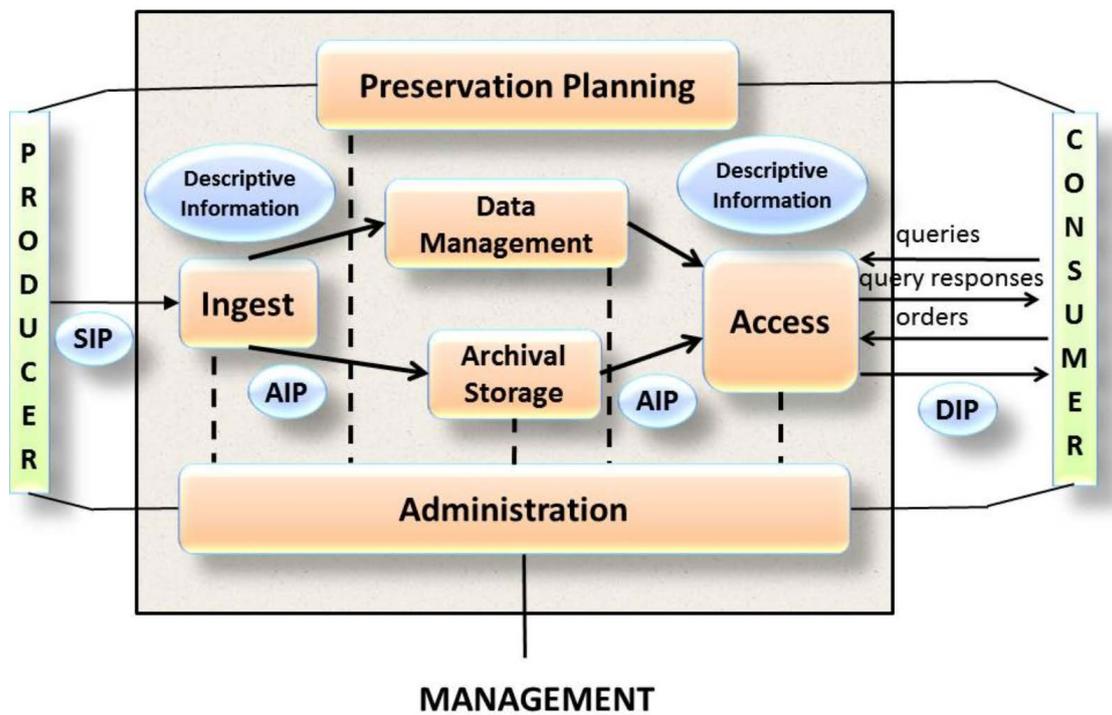


Abb. 1: OAIS-Referenzmodell¹⁹

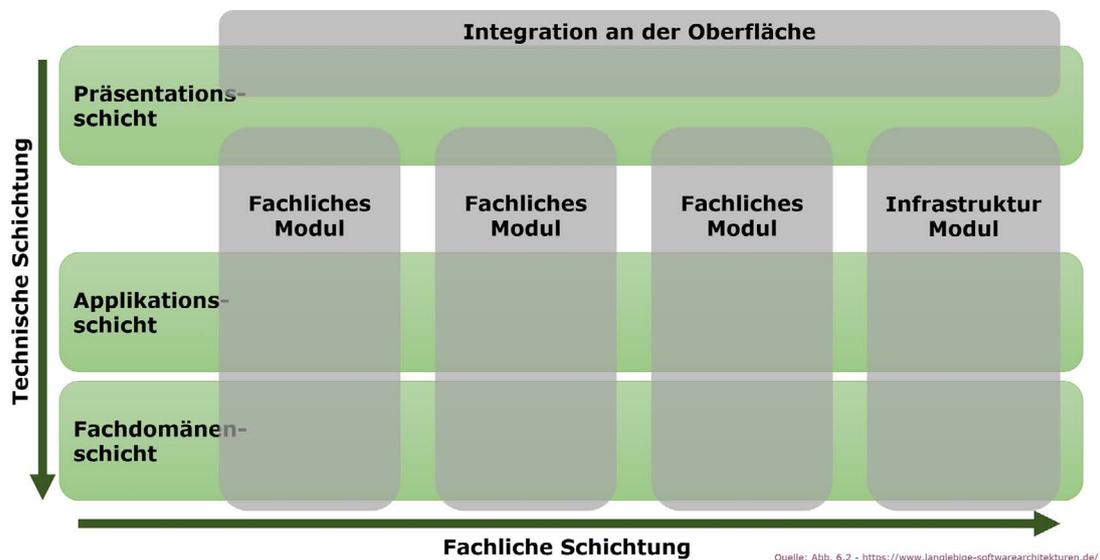


Abb. 2: Architektur-Modell (Modul steht synonym für Komponente)²⁰

19 Online unter: <http://www.digitalpreserve.info/wp-content/uploads/sites/2/2014/10/Slide5.jpg> (abgerufen am: 14.07.2023).

20 Carola Lilienthal, Langlebige Software-Architekturen: Technische Schulden analysieren, begrenzen und abbauen, Taschenbuch, Heidelberg 2017, Abbildung 6.2.

2.3 Vom OAIS-Referenzmodell zum Schichten- und Komponenten-Modell

Carola Lilienthal stellt in „Langlebige Software-Architekturen“²¹ eine sehr intuitiv verständliche Vorlage für die Darstellung des Architekturvorschlages bereit. Die Vorlage basiert u. a. auf den Ausarbeitungen von E. Evans zum Domain-Driven Design²² und G. Starke zu effektiven Software-Architekturen²³. Die sich hinter dem Modell verbergenden Grundgedanken werden nachfolgend zusammengefasst.

Heutige Schichtenarchitekturen enthalten eine fachliche und eine technische Dimension. Höhere Schichten dürfen darunterliegende Schichten verwenden. Das verwendete Modell enthält die Präsentations-, Applikations- und Fachdomänen-Schichten. Jeder Schicht werden Aufgaben zugeordnet und damit eine semantische Anreicherung nach technischen Gesichtspunkten herbeigeführt. Die fachliche Schichtung ist ein probates Mittel zur Hierarchisierung und Aufteilung der Fachlichkeit. Jedes fachliche Modul enthält Anteile der technischen Schichten. Früher wurde eine Infrastrukturschicht als zusätzliche unterste technische Schicht modelliert, doch daraus ergab sich das Dilemma, dass nicht direkt darüber liegende Schichten diese auch nicht direkt verwenden dürfen. Daher wird die bisher übliche Infrastrukturschicht als eine schichtenübergreifende Komponente modelliert.

Resultat ist: „Höhere Schichten, sowohl in der technischen als auch in der fachlichen Dimension der Schichtung, dürfen die weiter untenliegenden Schichten verwenden. Ein klares und einheitliches Muster!“²⁴

Jede Komponente besitzt eine Präsentationsschicht, dennoch soll sich das Gesamtsystem dem Benutzer als eine homogene Einheit präsentieren. Diese Aufgabe übernimmt eine integrierende Benutzeroberfläche.

2.4 Ingest (SIP)-, Archival Storage (AIP)- und Access (DIP)-Komponenten

Die **fachlichen Komponenten**, Ingest, Archival Storage und Access, stellen Funktionen, gekapselt durch Web-Services, zur Ver- und Bearbeitung der SIPs, AIPs und DIPs bereit. Das OAIS-Referenzmodell benennt die dauerhaft aufzubewahrende Domänen-Entität AIP (Archival Information Package) und die temporär aufzubewahrenden Domänen-Entitäten SIP (Submission Information Package) und DIP (Dissemination Information Package). Der Archival Storage dient der dauerhaften Aufbewahrung der AIPs. Für die SIPs und DIPs werden Storages zur nicht dauerhaften Aufbewahrung benötigt. In der Praxis zeigte sich, dass sich die Anforderungen an einen SIP- bzw. DIP-Storage „eigentlich nur“ in Bezug auf die Aufbewahrungsfrist unterscheiden. Daher bietet es sich an, die Storages für SIPs, AIPs und DIPs nach den gleichen Prinzipien zu konzipieren, respektive zu designen. Auf SIP, AIP und DIP sind teilweise die gleichen Funktionen anzuwenden, z. B. Format-Validierung beim Ingest, Format-Validierung bei der Bestandserhaltung (Preservation). Diese Funktionen werden in die Infrastruktur-Komponente ausgelagert.

In den letzten Jahren hat sich der de facto Standard REST zur Kapselung von Funktionalitäten auf Daten durchgesetzt und daraus schlussfolgernd und dem OAIS-Referenzmodell folgend, werden die drei **Software-Komponenten Ingest, Archival Storage und Access** gebildet, auf die nur über REST-Webservices zugegriffen werden kann. SIP, AIP und DIP können mittels Universally Unique Identifier (UUID)²⁵ eindeutig identifiziert und über Uniform Resource Identifier (URI)²⁶ eindeutig lokalisiert werden. Die Dokumentation der Web-Services wird durch den Einsatz eines OpenAPI²⁷-unterstützenden Frameworks erzeugt und kann archiviert werden.

21 Carola Lilienthal, wie Anm. 20.

22 Online unter: <https://leanpub.com/ddd-referenz/read> (abgerufen am: 14.07.2023).

23 Online unter: <https://www.gernotstarke.de/buecher#> (abgerufen am: 14.07.2023).

24 Lilithal, Langlebige Software-Architekturen, wie Anm. 20, S. 99.

25 Online unter: https://en.wikipedia.org/wiki/Universally_unique_identifier (abgerufen am: 14.07.2023).

26 Online unter: https://en.wikipedia.org/wiki/Uniform_Resource_Identifier (abgerufen am: 14.07.2023).

27 Online unter: <https://www.openapis.org> (abgerufen am: 14.07.2023).

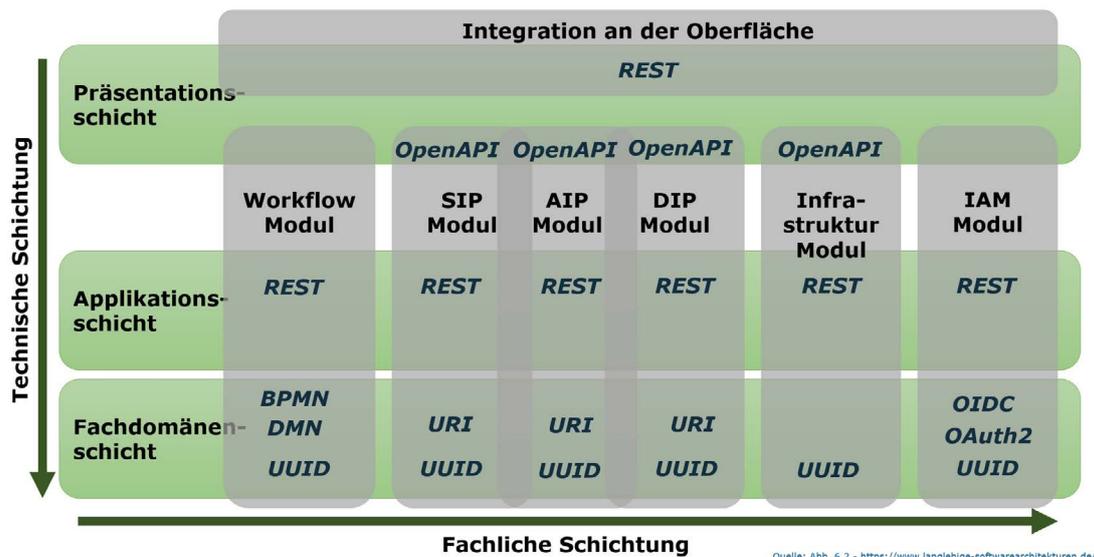


Abb. 3: Architektur-Modell und Standards

2.5 Workflow-Komponente

Die **Workflow-Komponente** verknüpft die von den fachlichen Komponenten, Ingest, Archival Storage und Access bereitgestellten Funktionen zu Prozessen und arbeitet diese ab.

Mit dem OMG-Standard BPMN 2.0¹⁰ lassen sich Workflows grafisch modellieren. Pools, Lanes, Tasks, Sub Prozesse, Gateways und Ereignisse bieten die größtmögliche Flexibilität bei der Modellierung.

- Pools identifizieren unterschiedliche Organisationen.
- Lanes beschreiben unterschiedliche Abteilungen.
- Tasks und Subprozesse können auch mehrfach parallel oder sequenziell abgearbeitet werden.
- Gateways beeinflussen die Ablaufsteuerung.

BPMN bietet die Task-Typen Receive Task, Send Task, User Task, Manual Task, Service Task, Script Task und Business Rule Task an.

Dem Workflow können beim Start Variablen übergeben werden, die während des gesamten Workflows benutzt und geändert werden können. Der Workflow kann während der Verarbeitung mit weiteren Variablen angereichert werden. Variablen dienen nicht nur der Prozesssteuerung, sondern können auch Process Meta Data sein.

Hinter Business Rule Tasks verbergen sich Decision Model and Notation (DMN)-Modelle²⁸ und diese entlasten Entwickler von der Entscheidungstabellen-Programmierung. Script Tasks werden vorrangig für die automatische Verarbeitung und Bearbeitung von Variablen eingesetzt. Service Tasks kapseln programmierte Geschäftslogik und bieten sich für die Verarbeitung von Web-Service Operationen an. BPMN-Modelle werden im XML-Format abgespeichert und sind selbstdokumentierend. Die Modelle können versioniert und unverändert archiviert werden. Eine BPMN-Engine kann BPMN-Modelle verarbeiten. Natürlich müssen die BPMN-Modelle um technische Parameter und Skript Snippets erweitert werden.

Das BPMN-Modell erarbeiten Fachbereich und IT gemeinsam und verfeinern es sukzessiv in einem iterativen Prozess. Das BPMN-Modell kann in eine BPMN-Engine deployed werden. Neue fachliche Anforderungen und Erfahrungen aus dem operativen Prozess führen zu weiteren Modelländerungen. Dieses Vorgehen ermöglicht eine stetige Verbesserung der Prozesse Softwareentwicklung und Administration, ganz im Sinne von software development (Dev) and information technology operations (Ops) (DevOps)²⁹.

28 Online unter: <https://www.omg.org/spec/DMN/About-DMN/> (abgerufen am: 14.07.2023).

29 Online unter: <https://en.wikipedia.org/wiki/DevOps> (abgerufen am: 14.07.2023).

2.6 Identity and Access Management³⁰-Komponente

Die **Identity Access Management-Komponente** umfasst alle Funktionalitäten zur Authentifizierung und Autorisierung. **Authentisierung** (englisch „authentication“) bezeichnet den Nachweis oder die Überprüfung der Authentizität. Die Authentisierung einer Identität kann u. a. durch Passwort-Eingabe, Chipkarte oder Biometrie erfolgen, die Authentisierung von Daten z. B. durch kryptographische Signaturen³¹. *Oder kurz formuliert:* Eingabe von Login-Daten in einem EDV-System (Behauptung einer Identität)³². **Authentizität** bezeichnet die Eigenschaft, die gewährleistet, dass ein Kommunikationspartner tatsächlich derjenige ist, der er vorgibt zu sein. Bei authentischen Informationen ist sichergestellt, dass sie von der angegebenen Quelle erstellt wurden. Der Begriff wird nicht nur verwendet, wenn die Identität von Personen geprüft wird, sondern auch bei IT-Komponenten oder Anwendungen. *Oder kurz formuliert:* Überprüfung der Behauptung durch das EDV-System inkl. Ergebnis der Prüfung (Verifizierung der Behauptung). Ein Anwendungssystem führt nach der **Authentisierung** durch den Benutzer eine **Authentifizierung** durch, indem es prüft, ob der Benutzername und das Passwort mit entsprechenden gespeicherten Daten übereinstimmen. Nach der Authentisierung, d. h. nachdem der Benutzer glaubwürdig seine Identität angegeben hat, folgt die Authentifizierung, d. h. die Überprüfung der Echtheit der Identität. Im Englischen gibt es dafür nur den Begriff **Authentication**³³. **Autorisierung** bezeichnet die Prüfung, ob eine Person, IT-Komponente oder Anwendung zur Durchführung einer bestimmten Aktion berechtigt ist. Nach der Prüfung wird ein Recht eingeräumt oder verweigert. Bestimmte Zugriffsrechte werden in Rollen zusammengefasst und diese Rollen werden Benutzern zugeordnet. Somit wird bei jedem Softwaresystem ein **Rechte- und Rollenkonzept** benötigt. Nachfolgend wird nur noch zwischen den Begriffen **Authentication (Authentifizierung)** und **Authorization (Autorisierung)** unterschieden.

Single Sign-on bedeutet, dass Nutzer/innen nach der einmaligen Authentifizierung an einem Arbeitsplatz (bei Web-Anwendungen ist der Browser gemeint) auf alle Dienste zugreifen können, für die sie autorisiert wurden. Wechselt die Person den Arbeitsplatz (Browser), verfallen Authentifizierung und Autorisierung. Im Frontend-Bereich und beim direkten Zugriff auf Backend-Services kann LDAP³⁴, ein verbreitetes Protokoll für den Zugriff auf verteilte Verzeichnisinformationen, zur Authentifizierung und Autorisierung genutzt werden. Die Kommunikation innerhalb des Backend-Bereiches kann mit LDAP nicht abgesichert werden. Single Sign-on kann im Frontend-Bereich durch die Verknüpfung von Kerberos³⁵ mit LDAP realisiert werden. Bei verteilten Systemen kann mit diesem Lösungsansatz kein Rechte- und Rollenkonzept vollständig umgesetzt werden. Der Standard Open ID Connect (OIDC) dient der Authentifizierung und OAuth2, Industrie-Standard-Protokoll³⁶, der Autorisierung. Beide Standards ermöglichen die Umsetzung eines Rechte- und Rollenkonzepts bei verteilten Anwendungen im Frontend- und Backend-Bereich. Open ID Connect wird von der OpenID Foundation³⁷ überwacht.

2.7 Infrastruktur- Komponente

Die **Infrastruktur-Komponente** fasst Funktionalitäten zusammen, die von allen anderen Komponenten benötigt und benutzt werden, und dient vorrangig der Vermeidung von Mehrfachimplementierungen z. B. Logging, XML-Validierung, Virenprüfung. Alle Funktionalitäten werden über REST-Webservices bereitgestellt.

2.8 Integrierende Benutzeroberfläche – User Interface

Die Benutzeroberfläche (engl. User Interface) ist die Komponente, die das Gesamtsystem für den Nutzer greifbar erscheinen lässt, und unterliegt eher technologischen und ergonomischen Änderungen im Gegensatz zu allen anderen bisher dargestellten Komponenten. Die Benutzeroberfläche soll alle vom Gesamtsystem bereitgestellten Funktionalitäten als eine homogene Einheit erscheinen lassen. Sie wirkt damit integrierend und wird als **integrierende Benutzeroberfläche** bezeichnet.

30 Online unter: https://de.wikipedia.org/wiki/Identit%C3%A4tsmanagement#Identit.C3.A4tsmanagement_von_Unternehmen (abgerufen am: 14.07.2023).

31 Online unter: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Kompodium/IT_Grundschutz_Kompodium_Edition2022.html (abgerufen am: 14.07.2023).

32 Online unter: <https://www.bsigroup.com/de-DE/our-services/digital-trust/cybersecurity-information-resilience/technology-solutions/cloud-security-solutions/identity--access-management> (abgerufen am: 14.07.2023).

33 Online unter: https://link.springer.com/chapter/10.1007%2F978-3-8274-2246-0_12 (abgerufen am: 14.07.2023).

34 Online unter: https://de.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol (abgerufen am: 14.07.2023).

35 Online unter: [https://de.wikipedia.org/wiki/Kerberos_\(Informatik\)](https://de.wikipedia.org/wiki/Kerberos_(Informatik)) (abgerufen am: 14.07.2023).

36 Online unter: <https://tools.ietf.org/html/rfc6749> (abgerufen am: 14.07.2023).

37 Online unter: <https://openid.net/foundation/> (abgerufen am: 14.07.2023).

2.9 Beschreibung der Schnittstellen

Alle Komponenten veröffentlichen REST³⁸-Webservice-API. REST selbst ist ein Architekturstil bzw. Programmierparadigma für verteilte Systeme. Der Autor fordert nicht zwingend, dass alle REST-Webservices zu RESTful-Webservices ausgebaut werden und damit das 1994 von Roy Fielding³⁹ entworfene Entwicklungsprinzip „Hypermedia as the Engine of Application State (HATEOAS)“ unterstützen. Diese Empfehlung gilt auch für die Anwendung von Hypertext Application Language (HAL)⁴⁰ auf JSON⁴¹ und XML⁴².

2.10 Dynamisches Verhalten – Logical View

Dieser Abschnitt spezifiziert das Verhalten des Systems im Detail. Der Abschnitt „Anwendungsfälle“ konzentriert sich auf die Spezifikation exemplarischer Anwendungsfälle des Systems aus externer sowie interner Sicht. Der Abschnitt Mechanismen liefert wichtige Modelle des systeminternen Verhaltens, die nicht durch den Abschnitt Szenarien abgedeckt werden.

Jeder Anwendungsfall wird optional dokumentiert

- durch eine Spezifikation, die definiert, wie sich die Architektur bei dem ausgewählten Anwendungsfall verhält und
- durch ein Komponenten-Interaktionsmodell, das beschreibt, wie die Komponenten zusammenwirken, um das gewünschte Anwendungfallergebnis zu erreichen.

2.11 Anwendungsfälle

Die beschriebenen Anwendungsfälle sollen die Möglichkeiten des Gesamtsystems verdeutlichen, beschränken sich aber auf die rein technischen Abläufe und berücksichtigen keine fachlichen oder organisatorischen Regelungen.

2.11.1 Anwendungsfall – REST-Webservice bereitstellen

Anwendungsfall	REST-Webservice veröffentlichen
Beschreibung	REST-Webservice wird für die rollenbasierte Benutzung veröffentlicht
Beteiligte	IAM-Administrator Softwareentwicklung
Komponenten	IAM, Ingest-, Archival-Storage-, Access- oder Infrastruktur-Komponente
Arbeitsschritte	1. REST-Webservice inkl. Open API-Dokumentation bereitstellen 2. REST-Webservice als OAuth2-Client veröffentlichen 3. Rolle dem OAuth2-Client zuordnen
Variationen	REST-Webservice für mehrere Realms

2.11.2 Anwendungsfall – Account eines Archivars einrichten (Ingest) – mit User Federation

Anwendungsfall	Account für Archivar einrichten
Beschreibung	Active Directory Account eines Archivars wird einer Rolle zugeordnet
Beteiligte	Identity and Access Management-Administrator (IAM) IT-Administration
Komponenten	IAM-Komponente
Arbeitsschritte	1. IAM-Administrator weist eine einer Person zugewiesene Active Directory-Gruppe einer entsprechenden Realm-Rolle zu
Variationen	Account soll Zugriff in mehrere Realms erhalten

2.11.3 Anwendungsfall – Login eines Benutzers (Access)

Anwendungsfall	Consumer-Login
Beschreibung	Ein Consumer kann sich mit einem seiner Social Login – Accounts anmelden
Beteiligte	IAM-Administrator
Komponenten	IAM-Komponente
Arbeitsschritte	1. IAM-Administrator muss einmalig für einen Realm einen Social Login Provider z. B. Google oder Facebook einrichten.
Variationen	Provider müssen je Realm eingerichtet werden

38 Online unter: https://de.wikipedia.org/wiki/Representational_State_Transfer (abgerufen am: 14.07.2023).

39 Online unter: https://de.wikipedia.org/wiki/Roy_Fielding (abgerufen am: 14.07.2023).

40 Online unter: https://en.wikipedia.org/wiki/Hypertext_Application_Language (abgerufen am: 14.07.2023).

41 Online unter: https://de.wikipedia.org/wiki/JavaScript_Object_Notation (abgerufen am: 14.07.2023).

42 Online unter: https://de.wikipedia.org/wiki/Extensible_Markup_Language (abgerufen am: 14.07.2023).

2.11.4 Anwendungsfall – Workflow entwerfen, veröffentlichen und abarbeiten

Anwendungsfall	Workflow entwerfen, veröffentlichen und abarbeiten
Beschreibung	Mit dem Camunda Modeler wird ein BPMN-Modell entworfen und Service-Tasks mit REST-Webservices umgesetzt
Beteiligte	Archivar, Informatiker
Komponenten	alle Komponenten
Arbeitsschritte	1. REST-Webservice Swagger-UI-Tool öffnen 2. Service Task konfigurieren 3. Workflow veröffentlichen 4. Workflow abarbeiten
Variationen	Abarbeitung nur mit bestimmten Berechtigungen

2.11.5 Anwendungsfall – E-Mail-Send-Task modellieren

Anwendungsfall	Send-Task konfigurieren
Beschreibung	Mit dem Camunda Modeler wird eine BPMN-Send-Task modelliert
Beteiligte	Archivar
Komponenten	alle Komponenten
Arbeitsschritte	1. E-Mail-Parameter konfigurieren from, to, subject und message-body. Der E-Mail Provider z. B. Exchange wird pro BPMN-Engine konfiguriert. 2. Workflow veröffentlichen 3. Workflow abarbeiten
Variationen	keine

Wichtige Aspekte zum Systemverhalten und zur Komponentenkollaboration werden in den Mechanismen erläutert, die im Abschnitt „Anwendungsfälle“ ausgelassen wurden. Auf diese kann an dieser Stelle nicht weiter eingegangen werden. Grundsätzlich wird beschrieben,

- wie die Komponenten zur Erfüllung der nicht-funktionalen Qualitäten und Einschränkungen zusammenarbeiten,
- welche Architekturstile und -muster benutzt werden,
- wie die Komponenten zusammenarbeiten und
- wie sich wiederholende Interaktionsmuster aussehen.

3 Other Views

Die anderen Sichten werden benutzt, um das System aus dem Blickwinkel der verschiedenen Stakeholder, wie Endnutzer, Entwickler oder Projektmanager, zu beschreiben. Physical und Process View ergänzen einander.

3.1 Development View

„Development View“ (auch Implementation View) beschreibt das System vom Standpunkt eines Entwicklers und beschäftigt sich mit dem Softwaremanagement. Die Entwicklung muss sich nur noch der Gestaltung der „Integrierende Benutzeroberfläche – User Interface“ und den „Ingest (SIP)-, Archival Storage (AIP)- und Access (DIP)-Komponenten“ widmen. In Zusammenarbeit mit dem Fachbereich werden die fachlich motivierten Workflows vom IT-Bereich nach technischen Gesichtspunkten umgestaltet und mit technischen Details ergänzt, z. B. REST-Webservice-Konfiguration, und der Workflow-Komponente zur Abarbeitung übergeben.

3.2 Process View

„Process View“ beschäftigt sich mit den dynamischen Aspekten des Systems und verdeutlicht die Prozesse des Systems und wie diese hinsichtlich des Laufzeitverhaltens kommunizieren. Parallelität, Verteilung, Integration, Performance und Skalierbarkeit sollen beschrieben werden.⁴³

Die Mandantenfähigkeit kann unterschiedlich umgesetzt werden:

- Mandantenfähigkeit durch redundante Gesamtsysteme, damit vertikal und horizontal skalierbar.
- Mandantenfähigkeit über Orchestrierungssoftware, damit vertikal und horizontal skalierbar.

Jedem Mandanten wird im IAM ein Realm (Domain) zugeordnet. Da jeder Web-Container-Anwendung nur ein Realm

43 Online unter: https://de.wikipedia.org/wiki/4%2B1_Sichtenmodell (abgerufen am: 14.07.2023).

zugeordnet werden kann, müssen diese redundant aufgebaut werden; dies beeinflusst aber die Skalierbarkeit des Gesamtsystems nicht.

Datenbank-Schemata werden mandantenbezogen zugeordnet; eine Datenbank kann 1 bis n Schemata bereitstellen. Die physische Trennung kann zwar über separate Datenbanken hergestellt werden, ist aber eher unüblich, wenn keine explizite physische Trennung, z. B. durch den umzusetzenden Schutzbedarf, verlangt wird. Alle sich hinter den Funktionseinheiten Ingest, Preservation Planning, Data Management, Archival Storage, Administration und Access verbergenden Prozesse können mit BPMN modelliert und von der Workflow-Komponente umgesetzt werden.

3.3 Physical View (Betriebssicht oder Laufzeitsicht)

„Physical View“ (oder auch deployment view) beschreibt das System vom Standpunkt des Systemarchitekten und beschäftigt sich mit der Verteilung der Softwarekomponenten auf physikalischer Ebene (also der Zuordnung dieser zu Hardware-Teilen) und deren Kommunikationen untereinander.

3.4 Hardware-Basis (Container-Virtualisierung)

Das Gesamtsystem kann auf physischer oder virtueller Hardware (Servern) aufgebaut werden. Der Autor empfiehlt dabei, die Methode der Container-Virtualisierung einzusetzen, die vorteilhaft für die Software-Entwicklung ist und im produktiven Betrieb die Ressourcen besser ausnutzt.

Vertikale Skalierung bedeutet, die Ressourcen eines physischen oder virtuellen Servers zu erhöhen z. B. Speichererhöhung, Erhöhung der CPU-Anzahl. *Horizontale Skalierung* bedeutet, dass physische oder virtuelle Server hinzugefügt werden. Diese Skalierungen sind mit physischen oder virtuellen Servern nur aufwendig realisierbar. Einen Ausweg bietet die Container-Virtualisierung deren Container mittels Orchestrierungssoftware „leichtgewichtig“ horizontal und vertikal skaliert werden können.

Als Software zur Container-Virtualisierung hat sich Docker⁴⁴ und zur Orchestrierung Kubernetes⁴⁵ durchgesetzt. Die IT-Dienstleister werden den Archiven eine BSI-zertifizierte Lösung anbieten.

3.5 Identity and Access Management-Komponente

Die „Identity and Access Management-Komponente“ kann mit der Community Software Keycloak⁴⁶ realisiert werden. (Hinweis: Seit März 2018 ist Keycloak als JBoss-Community-Projekt unter der Leitung von Red Hat⁴⁷ verfügbar, die es als Upstream-Projekt für ihr RH-SSO-Produkt verwenden.)

Keycloak ist eine JBoss/Wildfly-Web-Container-Anwendung, basierend auf Java 8 JDK⁴⁸ und benötigt im produktiven Betrieb eine SQL⁴⁹-Datenbank, die über JDBC⁵⁰ angebunden wird⁵¹.

Keycloak bietet:

- eine rollenbasierte Autorisierung,
- Single Sign-on,
- User Federation (LDAP, Active Directory),
- Social Login⁵²,
- Client-Adapter⁵³ für viele Plattformen und Programmiersprachen,
- eine Account Management Console, über die Nutzer ihren Account pflegen können,
- eine Administration Console,
- ein Administrator Command Line Interface⁵⁴ und
- ein Administrator REST API⁵⁵.

44 Online unter: <https://www.docker.com/> (abgerufen am: 14.07.2023).

45 Online unter: <https://kubernetes.io> (abgerufen am: 14.07.2023).

46 Online unter: <https://www.keycloak.org/> (abgerufen am: 14.07.2023).

47 Online unter: https://de.wikipedia.org/wiki/Red_Hat (abgerufen am: 14.07.2023).

48 Online unter: https://de.wikipedia.org/wiki/Java_Development_Kit (abgerufen am: 14.07.2023).

49 Online unter: <https://de.wikipedia.org/wiki/SQL> (abgerufen am: 14.07.2023).

50 Online unter: https://de.wikipedia.org/wiki/Java_Database_Connectivity (abgerufen am: 14.07.2023).

51 Informationen zum Datenbanksetup können hier nachgelesen werden: <https://www.keycloak.org/server/db> (abgerufen am: 14.07.2023).

52 Online unter: https://en.wikipedia.org/wiki/Social_login (abgerufen am: 14.07.2023).

53 Online unter: https://www.keycloak.org/docs/latest/securing_apps/#supported-platforms (abgerufen am: 14.07.2023).

54 Online unter: https://www.keycloak.org/docs/latest/server_admin/index.html#the-admin-cli (abgerufen am: 14.07.2023).

55 Online unter: <https://www.keycloak.org/docs-api/19.0.3/rest-api/index.html> (abgerufen am: 14.07.2023).

Keycloak unterstützt

- die Authentifizierungsprotokolle OpenID Connect und SAML⁵⁶ 2.0, wobei OpenID Connect bevorzugt wird, und
- das Autorisierungsprotokoll OAuth 2.0.

„User Federation“ wird vorzugsweise für die archivinterne Authentifizierung und Autorisierung und „Social Login“ für die der Consumer eingesetzt. Social Login erspart den Archiven die Verwaltung von Consumer-Accounts.

3.6 Workflow-Komponente

Die „Workflow-Komponente“ kann mit der Open-Source-Software der Firma Camunda⁵⁷ realisiert werden. Notwendig sind die Produkte⁵⁸:

- BPMN Engine
- DMN Engine
- Modeler (Entwicklungswerkzeug)
- Tasklist (Nutzertaskverwaltung)
- Cockpit (Technischer Betrieb)
- Admin (Nutzer- und Rechteverwaltung)

Bis auf den Modeler sind alle Produkte Java-Web-Container-Anwendungen⁵⁹, die für folgende Java-EE-Application-Server (mind. Java 8⁶⁰) unterstützt werden:

- Tomcat⁶¹
- JBoss⁶²/Wildfly⁶³
- IBM WebSphere⁶⁴
- Oracle WebLogic⁶⁵

Der Modeler wird als standalone- oder web-basierte-Anwendung bereitgestellt. Bis auf den Modeler können alle Produkte über eine REST API gesteuert werden. Single Sign-on ist sowohl mit Keycloak als auch mit Kerberos in Verbindung mit LADP⁶⁶ oder dem Active Directory möglich. Camunda stellt mit dem Produkt „Admin“ eine eigene Nutzer- und Rechteverwaltung bereit, die mit Keycloak verknüpft werden kann.

Die Camunda-Software-Produkte benötigen eine Datenbank; folgende Produkte⁶⁷ werden von Camunda unterstützt (Stand: 07.01.2020):

- MySQL⁶⁸ 5.6 / 5.7
- MariaDB⁶⁹ 10.0 / 10.2 / 10.3
- Oracle⁷⁰ 11g / 12c / 18c
- IBM DB2⁷¹ 10.5 / 11.1 (Mit Ausnahme von z/OS in allen Versionen)
- PostgreSQL⁷² 9.4 / 9.6 / 10.4 / 10.7 / 11.1 / 11.2

56 Online unter: https://de.wikipedia.org/wiki/Security_Assertion_Markup_Language (abgerufen am: 14.07.2023).

57 Online unter: <https://camunda.com/de/> (abgerufen am: 14.07.2023).

58 Online unter: <https://camunda.com/de/enterprise/> (abgerufen am: 14.07.2023).

59 Online unter: <https://docs.camunda.org/manual/latest/installation/full/> (abgerufen am: 14.07.2023).

60 Online unter: <https://docs.camunda.org/manual/latest/introduction/supported-environments/#java> (abgerufen am: 14.07.2023).

61 Online unter: <https://tomcat.apache.org/> (abgerufen am: 14.07.2023).

62 Online unter: <https://www.jboss.org/> (abgerufen am: 14.07.2023).

63 Online unter: <https://www.wildfly.org/> (abgerufen am: 14.07.2023).

64 Online unter: <https://www.ibm.com/cloud/websphere-application-platform/> (abgerufen am: 14.07.2023).

65 Online unter: <https://www.oracle.com/de/middleware/weblogic/> (abgerufen am: 14.07.2023).

66 Online unter: <https://docs.camunda.org/manual/latest/installation/standalone-webapplication/#ldap-configuration> (abgerufen am: 14.07.2023).

67 Online unter: <https://docs.camunda.org/manual/latest/introduction/supported-environments/#databases> (abgerufen am: 14.07.2023).

68 Online unter: <https://www.mysql.com/de/> (abgerufen am: 14.07.2023).

69 Online unter: <https://mariadb.org/> (abgerufen am: 14.07.2023).

70 Online unter: <https://www.oracle.com/database/> (abgerufen am: 14.07.2023).

71 Online unter: <https://www.ibm.com/analytics/db2> (abgerufen am: 14.07.2023).

72 Online unter: <https://www.postgresql.org/> (abgerufen am: 14.07.2023).

- Amazon Aurora⁷³ PostgreSQL kompatibel mit PostgreSQL 9.6 / 10.4 / 10.7
- Microsoft SQL Server⁷⁴ 2012/2014/2016/2017
- H2 1.4⁷⁵

Die zu veröffentlichenden (deployed) BPMN-Modelle (Workflows) werden vom Modeller oder per cURL⁷⁶ an die BPMN- bzw. DMN-Engine als XML-Datei übergeben und sind damit archivierbar.

3.7 Integrierende Benutzeroberfläche – User Interface

Die „Integrierende Benutzeroberfläche – User Interface“ kann mit der „low-code“ Entwicklungsplattform Oracle Application Express (APEX)⁷⁷ der Firma Oracle realisiert werden. Lizenzrechtlich ist es ein „no cost database feature“. Oracle APEX bietet sich als Entwicklungsplattform für die Umsetzung der integrierenden Benutzeroberfläche an. Es selbst ist eine Oracle-Datenbank-Anwendung, die von jeder Oracle Datenbank Edition ab Version 11 unterstützt wird. Oracle APEX benötigt neben der Oracle Datenbank die Java-Container-Anwendung Oracle REST Data Service (ORDS)⁷⁸, wobei mit ORDS zusätzlich u. a. RESTful-Webservices bereitgestellt werden können. ORDS kann auf folgenden Java-EE-Application-Servern bereitgestellt werden:

- Oracle Web Logic⁷⁹
- Tomcat⁸⁰

Single Sign-on ist sowohl mit Keycloak als auch mit Kerberos in Verbindung mit LDAP⁸¹ oder dem Active Directory möglich.

3.8 Infrastruktur-Komponente

In die „Infrastruktur- Komponente“ werden alle Funktionen aufgenommen, die von allen fachlich darüber liegenden Komponenten gemeinsam benutzt werden. Alle Funktionen werden als REST-Webservices bereitgestellt. Beispielsweise wird das Logging als wiederverwendbarer REST-Webservice der Infrastruktur-Komponente zugeordnet. Es gelten die „Allgemeine Hinweise zur Entwicklung von REST-Webservices“.

3.9 Ingest (SIP)-, Archival Storage (AIP)- und Access (DIP)-Komponenten

Die „Ingest (SIP)-, Archival Storage (AIP)- und Access (DIP)-Komponenten“ kapseln die Datenbenutzung und alle Be- und Verarbeitungsfunktionen über REST-Webservices. Jedes SIP, AIP und DIP wird über eine UUID referenziert und kann über eine URI lokalisiert werden. REST-Webservices werden in die Infrastruktur-Komponente ausgelagert, wenn diese auch von anderen Komponenten benötigt werden. Jede Komponente deckt die Fachdomänen-, Applikations- und Präsentationsschichten ab und ermöglicht damit eine autarke Benutzung. Der Fachdomänenschicht werden die Fachdomänen-Entitäten SIP, AIP und DIP und alle Be- und Verarbeitungsfunktionen zugeordnet. Der Applikationsschicht werden die Funktionen kapselnden REST-Webservices zugeordnet. Der Präsentationsschicht werden Web-Anwendungen und die mit Swagger generierten UI-Tools zugeordnet. Es gelten die „Allgemeinen Hinweise zur Entwicklung von REST-Webservices“.

3.10 Allgemeine Support-Hinweise für alle Komponenten

Alle oben aufgeführten Produkte sind vertikal und horizontal skalierbar. Die Hersteller bieten bei Bedarf einen kostenpflichtigen professionellen Support an.⁸²

73 Online unter: <https://aws.amazon.com/de/rds/aurora/> (abgerufen am: 14.07.2023).

74 Online unter: <https://www.microsoft.com/de-de/sql-server/sql-server-downloads> (abgerufen am: 14.07.2023).

75 Online unter: <http://h2database.com/html/main.html>. Bezüglich des Clusterings siehe die Angaben von Camunda; online unter: <https://docs.camunda.org/manual/latest/introduction/supported-environments/#database-clustering-replication> (abgerufen am: 14.07.2023).

76 Online unter: <https://curl.haxx.se/> (abgerufen am: 14.07.2023).

77 Online unter: <https://apex.oracle.com/de/> (abgerufen am: 14.07.2023).

78 Online unter: <https://www.oracle.com/database/technologies/appdev/rest.html> (abgerufen am: 14.07.2023).

79 Online unter: <https://docs.oracle.com/en/database/oracle/oracle-rest-data-services/19.2/aelig/installing-REST-data-services.html#GUID-63AD5B59-CD6A-47A6-BBB4-0D614090C51F> (abgerufen am: 14.07.2023).

80 Online unter: <https://docs.oracle.com/en/database/oracle/oracle-rest-data-services/19.2/aelig/installing-REST-data-services.html#GUID-3F2AE730-69D0-4A64-A13A-76745B7467CD> (abgerufen am: 14.07.2023).

81 Online unter: <https://docs.camunda.org/manual/latest/installation/standalone-webapplication/#ldap-configuration> (abgerufen am: 14.07.2023).

82 Siehe hierzu die weiterführenden Hinweise auf den Homepages von Camunda, Keycloak, Oracle APEX, HA Proxy, Tomcat Apache usw.

3.11 Allgemeine Hinweise zur Entwicklung von REST-Webservices

Alle selbst entwickelten REST-Webservices sind mit OpenAPI zu dokumentieren. Das Open-Source-Framework Swagger⁸³ unterstützt diese API Beschreibungssprache. Die mit Swagger generierten REST-Webservice-Dokumentationen sind archivierbar. Das Swagger-UI-Tool bietet wesentliche Vorteile für Softwareentwickler bei der Softwareentwicklung und dem Softwaretest. Die Benutzung der REST-Webservices wird über das Identity and Access Management abgesichert und erfolgt immer rollenbasiert. Es muss ein Konzept für die Versionierung und Dokumentation von REST-Webservices erarbeitet werden.

3.12 Allgemeine technische Hinweise für alle Komponenten

Bis auf die Komponente „Integrierende Benutzeroberfläche – User Interface“ deckt jede Komponente die Fachdomänen-, Applikations- und Präsentationsschichten ab und ermöglicht eine autarke Komponentenentwicklung und Abgrenzungen durch eindeutige Schnittstellen. Die horizontale Skalierbarkeit wird unterstützt, wenn alle Web-Anwendungen je Komponente durch einen Proxy-Server (z. B. Apache HTTP-Server oder IIS von Microsoft) gekapselt werden. Stellt eine Komponente neben http bzw. https auf anderen Protokollen basierende Schnittstellen bereit, dann sollten alle Protokolle mit einem HAProxy⁸⁴ zusätzlich gekapselt werden. Dadurch wird eine hohe Flexibilität bei der Portkonfiguration einer Komponente erreicht.

3.13 Konzeptionelles Framework

In diesem Abschnitt werden alle bisherigen Konzepte und sonstigen Ausarbeitungen zusammengefasst und, sofern möglich, in Beziehung zueinander gestellt. Die Konzepte werden in einem Domänen-Lexikon dokumentiert. In das Domänen-Lexikon werden zusätzlich alle wichtigen Begriffe aus der Problemdomäne und aus den Diskussionen über das System und seiner Architektur aufgenommen. Weitere Ausführungen müssen an dieser Stelle unterbleiben.

4 Fazit

Schlussfolgerung: Dieser Software-Architektur-Vorschlag favorisiert den Einsatz standardumsetzender und markt-erprobter Produkte und schafft damit Freiräume für die fachliche Entwicklung. Standards ermöglichen eine breitere Produktauswahl und verringern Abhängigkeiten gegenüber Software-Herstellern. Die Bereiche Workflow und Identity and Access Management sind in jedem größeren Projekt relevant und sollten niemals selbst entwickelt werden. Ähnlich verhält es sich bei der integrierenden Oberfläche.

Vorteile und Grenzen: Der Autor kann aus seinen bisher mit dem bereits vorhandenen Framework gesammelten Erfahrungen berichten, dass fachliche Anforderungen in kürzester Zeit (mind. 1:10) umsetzbar sind. Besonders vorteilhaft ist es, Fachkomponenten sukzessiv entwickeln zu können, denn: Wer kann die fachlichen Anforderungen am Projektanfang umfassend beschreiben? Flexibilität und Erweiterbarkeit werden benötigt und mit BPMN geboten. Zukünftige Projektaufträge werden sich vorrangig der Entwicklung von REST-Webservices widmen, sind klar spezifizierbar und können gegebenenfalls extern beauftragt werden. Es kann auch die Umsetzung ganzer Workflows beauftragt werden. Der Auftragnehmer könnte entweder mit einer Kopie des out-of-the-box-Systems arbeiten oder erhält einen eigenen, durch einen Realm abgegrenzten Testbereich für die Auftragsdurchführung. Die IT-Dienstleister der Archive können bei der technischen Umsetzung der Gesamtarchitektur bei jeder Komponente zwischen Produkten unterschiedlicher Hersteller nach eigener Präferenz wählen. Die Archive werden feststellen, dass die Gesamtarchitektur auf kein „exotisches“ Produkt verweist und damit den IT-Dienstleister mit unüblichen Problemen konfrontiert.

- Jedes Archiv kann innerhalb seines Realms das eigene Rechte- und Rollenkonzept umsetzen.
- Nicht-funktionale Anforderungen: Alle Anforderungen werden erfüllt.
- Evtl. vorhandene Inkonsistenzen: Sind dem Autor nicht bekannt oder bewusst.
- Mögliche Richtungen für die Entwicklung der Architektur: Der Autor ist davon überzeugt, dass sich die Container-Virtualisierung durchsetzt, weil die ökonomischen Erfordernisse es gebieten (z. B. kürzere Entwicklungszeiten, flexiblere Software, bessere Ressourcen-Ausnutzung).

Die Bewertung der Erfüllung aller nicht-funktionalen Anforderungen sollte vom Archiv vor der Entscheidung für vorgeschlagene Software-Architektur durchgeführt werden.

83 Online unter: https://de.wikipedia.org/wiki/Swagger_%28Software%29 (abgerufen am: 14.07.2023). Online unter: <https://swagger.io/> (abgerufen am: 14.07.2023).

84 Online unter: <https://www.haproxy.org/> (abgerufen am: 14.07.2023).

Nach einer Entscheidung für diesen Software-Architektur-Vorschlag müssen:

- ein Rechte- und Rollenkonzept exemplarisch erarbeitet und sukzessiv fortgeführt werden,
- ein Konzept zur Versionierung und Dokumentation von REST-Webservices erarbeitet und sukzessiv fortgeführt werden und
- Schulungen in den Bereichen
BPMN (Camunda),
Keycloak (OAuth2, OpenID Connect),
REST-Webservices inkl. JSON (Java),
Oracle APEX,
Logging,
Docker und Kubernetes
durchgeführt werden.

Bisher liegt im DIMAG-Entwicklungsverbund kein Konzept zur Speicherschicht vor. Die Software-Architektur begegnet diesem Mangel mit der Kapselung der SIP, AIP und DIP über REST-Webservices. Durch die Identifizierung aller SIP, AIP und DIP mit UUIDs und die Lokalisierung über URIs besteht eine klare Trennung zwischen logischer Adressierung und physischer Speicherung.

4.1 Kosten

Der gesamte Software-Architektur-Vorschlag kann mit Open-Source- und Community-Produkten umgesetzt werden. Es entstehen nur zusätzliche Kosten, wenn ein professioneller Support benötigt wird. *siehe Allgemeine Support-Hinweise für alle Komponenten.*

Strategies in digital archiving at the Belgian State Archives. Between ideal and reality

Rolande Depoortere and Johan Van der Eycken, *Archives générales du Royaume, Bruxelles*

The large-scale computerization of government services took place from the '80s and '90s of the last century, but certainly accelerated after the millennium change with the introduction of new applications and technologies, an advancement whose peak has not yet been reached. This development has immediate consequences for archival institutions, which must prepare themselves to preserve these digital files for the future. Given the rapid development of technologies, file formats, media, etc., this is not a simple and therefore a costly undertaking. The risk of losing digital information is often greater than the risk of losing the same information on analogue media. In this article, we examine how the Belgian State Archives dealt with this new mission, beginning with the institutional and political context in which this strategy was defined. Next, we will address some concrete realizations to conclude with a SWOT analysis of the chosen path.

1 Strategy determined by the institutional and political context

As all federal institutions, the State Archives have been facing drastic budget cuts for several years. This is happening at a moment when the mission of acquiring, preserving and disseminating digital-born archives is added to the collection, preservation and dissemination of analogue documents, thus at a moment when the need for new financial resources, new competence profiles, new staff is growing. The political and institutional context makes any budgetary increase impossible. In 2014, we were confronted with two ageing basic tools that had to be replaced by new solutions: the archive management system *Archeion*, originally developed by the Dutch National Archives and maintained by a private Dutch company, and the search engine developed by another private Dutch firm. Both companies had no intention to extend the maintenance of these specific, tailor-made tools because there were too few clients. The continuation of these tools was no longer in line with their commercial strategy.

After more than one year of benchmarking and unsuccessful search for financial means necessary to externalize the development of software, to acquire or to lease new tools, we had the opportunity to engage two young web developers to replace retiring archivists and administrative collaborators. Note that no wage increase was foreseen in this case, only a redefinition of competence profiles and a redistribution of recruitment priorities. Our staff included already one web developer, an IT engineer, and a web designer/webmaster. With two more colleagues, a real development team could be set up. The development team is associated to a service composed of two archivists and two or three collaborators under fixed-term contracts that are paid through research projects with a term of one or maximum two years. Our annual operation budget is almost zero: just 50,000 euros, 100 % of which is devoted to paying the external hosting of our search engine and website. To make it short: there is no money for externalization, purchase, or leasing of tools.

So, these circumstances dictated our policy and forced us to opt for a strategy that follows five axes.

Axis 1. In-house development of open-source applications

The archivists of our service carry out an analysis of the user needs and functional requirements, while the IT experts oversee the actual development. We support the open-source philosophy, but we are obliged to bear in mind that the State Archives' working environment reposed on Windows. All applications must of course be compliant with the current international standards and norms, they must all be interoperable with the other components of our archival management and preservation system as well as with external platforms and systems. In this context, it is noteworthy that the State Archives is a member of different international networks like *Archives Portal Europe* (APE), the *Consortium of European Social Science Data Archives* (CESSDA ERIC) and the *Digital Research Infrastructure for the Arts and Humanities* (DARIAH ERIC).

Axis 2. Integration of existing tools in our archival ecosystem

We integrate existing tools and applications in our archival ecosystem under the condition that they are open source and freely available. Integration, implementation and customization to our specific needs and workflows must be realized by our own staff or with external support but, in the latter case, this support must be free of charge.

Axis 3. Externalization to partners for development and parametrization of applications

We try to conclude partnerships with other federal institutions, with universities or private companies that develop applications for us respectively carry out a functional or user needs analysis for us. These kinds of partnerships can be organized within the scope of innovative (research) projects financed by third parties. The federal ministry for Scientific

Policy, the Flemish funding agency for Scientific Research, its counterpart on the French-speaking side and some private benefactor foundations are the main financiers. This method has a major disadvantage: working within the framework of a project means that the availability of resources is limited in time and this timeline does not align well with the obsolescence of our infrastructure and software.

During the last five years, the State Archives has been a partner institution in six projects related to digital archiving:

- MADDLAIN (2015-2017) studied the needs and expectations of the digital users of our institution and of the Royal Library.
- HECTOR (2015-2018) analyzed the management of hybrid information produced in the federal public sector and formulated recommendations to improve business processes and introduce best practices.
- PROMISE (*Preserving Online Multiple Information: towards a Belgian StratEgy (2018–2019)*) built a prototype for archiving the “Belgian” Web. This project was coordinated by the Royal Library.
- SODA (*Social Sciences Data Archive*) & BISHOPS (*Belgian Infrastructure for Sociale Sciences and Humanities open Cloud*) are two linked projects that aim to build and develop an infrastructure for the preservation and dissemination of research data produced by Belgian research institutions (universities and federal research institutes) in social sciences and digital humanities. This infrastructure is the Belgian national node of CESSDA ERIC. At the present time, the infrastructure is hosted by the State Archives, who brings its expertise in long-term preservation to the project.
- DIGHIMAPS (*Digital Maps and Archives – Activating Cartographic Collections in a Digital World*): explores the means of enhancing the geo-localization and the extraction of location names from old handwritten maps.

Axis 4. Externalization of some functions to federal institutional partners

The fourth axis is the delegation of some of our legal missions to other federal operators that we assist with expertise in long-term archiving.

Axis 5. Promotion of records management and digital archiving in the public sector

Our last action line comprises promoting records management and digital archiving best practices in the public sector, to avoid the loss of digital information before data transfer to the State Archives. Through this proactive action we hope to make our job easier when digital records are transferred to us.

Infrastructure for digital archiving

The following chapter illustrates these five axes that determine the development of our digital archival system. Our system for digital archiving follows the OAIS model (ISO 14721:2012), which has become the standard for such systems. Figure 1 illustrates the position that each tool has in the system.

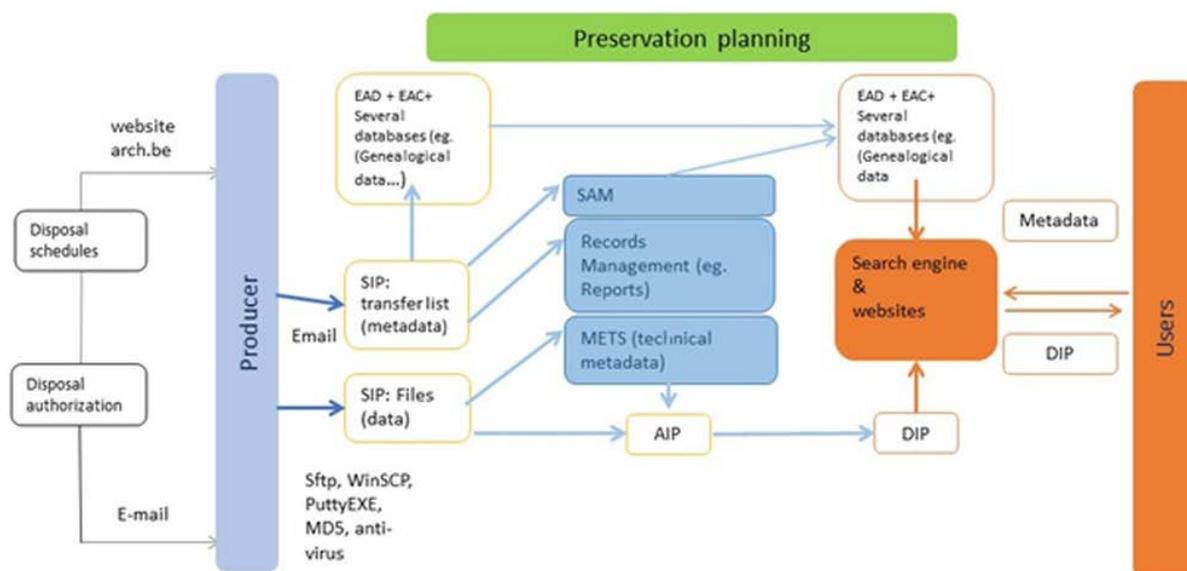


Figure 1: Flowchart of the State Archives' infrastructure for digital archiving according to the OAIS-model

2 In-house development of open-source applications

2.1 SAM: The State Archives Management application

The SAM-application manages the entire workflow of the State Archives for both the analog and digital holdings. Now, three modules are operating: the registration and management of archives collections, the registration of users, and the monitoring of access (orders, special authorizations, loan of documents, etc.). SAM makes what the old system did but better and, above all, it also processes digital-born archives and hybrid collections, which the old system could not.

We applied the AGILE method during development by bringing together a panel of individuals that represented internal and external users, and organized usability tests, which were followed by immediate minor or major adjustments in the code. SAM is developed in PHP with a MariaDB database and works with database servers and an Apache server. The interfaces of SAM are available in French, Dutch and German, the three official languages in Belgium (the State Archives also has a German-speaking service in Eupen). Special attention was devoted to user-friendliness, simplicity and a sober design.

The administrative and technical metadata (store location, extent, rights of access, etc.) of the digital and analog records are registered in SAM, and so are the general descriptive metadata at the archive collection level (title, archive creator, period, etc.). But the series and item level descriptions are not in SAM: they exist in *Encoded Archival Description*/xml files that are searchable and readable through the search engine. SAM has a link to the EAD/xml files and to our search engine to display the complete description of the series and items on the SAM interface.

Visitors residing in the European Union are registered with their electronic identity card. A specific module automatically extracts data from the e-ID card. This same module supports the creation of bills for fee-based services.

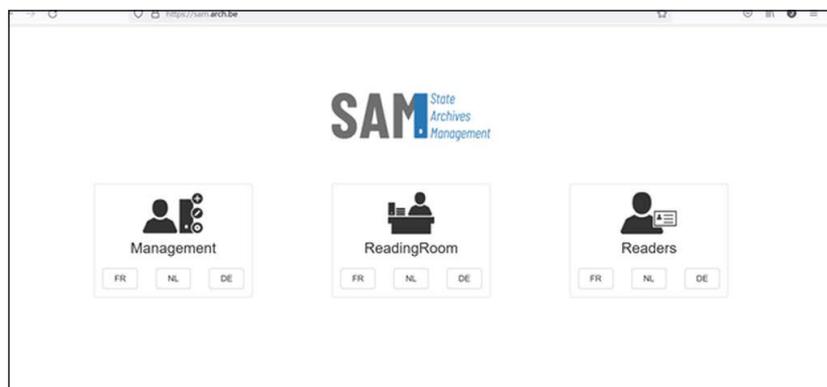


Figure 2 Start screen of the State Archives Management application

Additional modules will be developed in 2023 and 2024 to have a complete software suite that monitors the selection, transfer, registration of and access to digital-born and analog archives:

- The first one will support the delivery of authorizations to eliminate documents to records creators, i.e., the creation of disposal lists by records creators and their validation by the State Archives.
- The second module will support the creation of transfer lists by records creators and their validation by the State Archives, and then the transfer of digital-born archives in a fully integrated web application. At present, the transfer of digital files is made through WinSCP, which must be installed on the server of archive creators.
- The third module will link the SAM database to work documents, emails and documentation about archive collections and records creators.

2.2 The new search engine and website (2023)

The user expectation analysis of the MADDLAIN project staff consisted in an online survey. We collected about 1,000 answers. Most respondents were genealogists, so we completed the data with face-to-face interviews with researchers as well as a workshop that brought together representatives of Belgian universities: researchers, students, and teachers. A student trainee in library, archives and documentation sciences devised a new website and search engine design.

The development methodology will be the same as for SAM: AGILE method and usability tests with a panel of users. We have drawn user profiles and selected volunteers for the panel.

The replacement of the search engine will give the State Archives the opportunity to implement an open data policy and the FAIR principles as much as possible. Unlike today, visitors of our search website will have the right to download millions of images from our collections for free, but with a limit of 100 files per day and without the possibility of batch-downloading (to avoid overload and slowdown of the site).

3 Integration of existing tools in our archival ecosystem

3.1 Transfer of digital-born administrative documents

Presently, our infrastructure relies on SFTP and on the open source WinSCP client program, which archive creators must install on their server. WinSCP was chosen because 99 % of Belgian public services work with Windows. This software is also very user-friendly, with a drag-and-drop function to launch the transfer. We are testing other applications, especially those that allow treatment of large bulks of unstructured office documents. In 2020, we attended a demonstration of *OCTAVE* by the *Archives Nationales de France*. We also tested *Archifiltre*, and two software programs that can rename files in batch: *VRenamer* (too basic for our needs) and *Bulk Rename Utility* (more efficient). *Archifiltre* is very helpful: it can display a file tree, find duplicate files, and rename files to create *Submission Information Packages* (SIPs).

We closely follow the recommendations of the E-Ark project. The E-Ark tools will be tested soon: ERMS Export Module, RODA-ind, ETA, ETP and the Database Preservation Toolkit. Depending on the results, our plan to develop new SAM modules will be reevaluated.

3.2 Transfer, dissemination, and reuse of research data

Dataverse for research data: *Dataverse* is an open source, community-driven software platform developed by the Harvard University. It allows to integrate other data services such as *DataCite* (DOI provider) or *ROpenScience*. It has a modular design principle and relies on APIs that enable distributed file storage. It also enables the development of further micro-services. *Dataverse* has been chosen by several CESSDA service providers such as the Dutch and the Austrian. At the European level, *SSHOC (the European Social Sciences & Humanities Open Cloud)* is developing a data repository service for social sciences and humanities institutions, built upon the *Dataverse* software. *Dataverse* has also been selected by several Belgian universities for their local cross-disciplinary data repository.

We will keep using *Dataverse* for transfer and dissemination of research data because it is strongly oriented towards the research community. However, it is not really suited for transfer and dissemination of administrative archives. The installation and customization of *Dataverse* require strong IT competences. Despite a smooth and continuous dialogue with the Harvard developers, they cannot always prioritize user-driven adaptation requests, which means all our needs cannot be addressed as early as we wish.

Note that *Dataverse* is a transfer and dissemination platform but in no way an archival management program or a preservation tool. It is an intermediate layer on top of the data management system. We are planning to build a bridge between *Dataverse*, SAM, our own general search engine and the long-term preservation infrastructure. For instance, we had to devise a mapping between the EAD and the DDI (Data Documentation Initiative) metadata models. DDI-encoded metadata are a CESSDA requirement, and *Dataverse* can automatically format metadata in DDI CodeBook 2.5, among other metadata schemas.

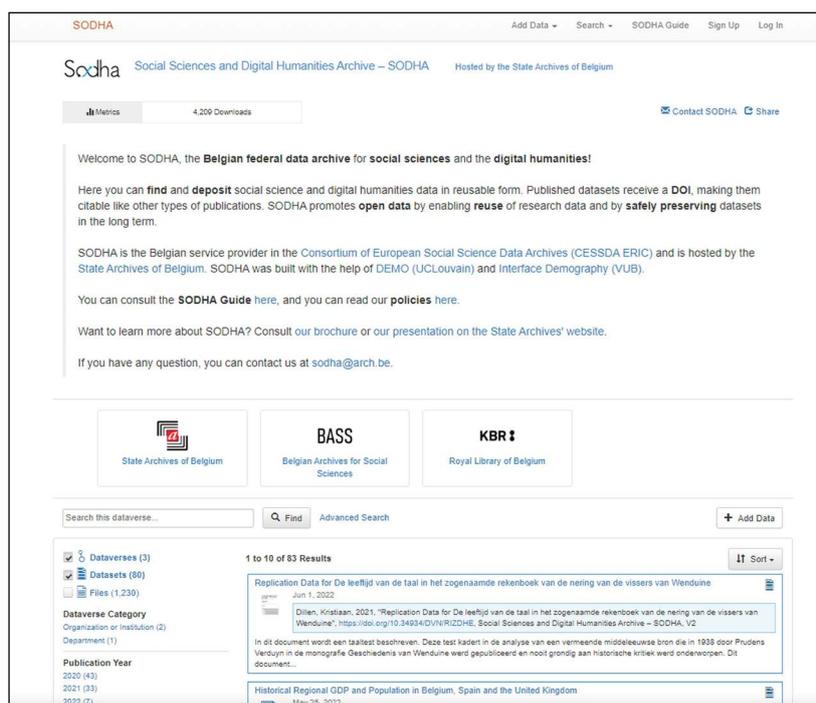


Figure 3 Interface of the Social Sciences and Humanities Archive

3.3 Web archiving

For Web archiving, the PROMISE project team selected a series of software solutions and combined them with some in-house development. Here is a short review of the selected tools: To carry out a selection of websites, we needed a tool that simplifies the creation of seed lists of websites, and which could be linked to the *Heritrix* configuration. This component is developed with Python, Django and PostgreSQL. It relies on the OCLC metadata model.

- For capturing websites, we use the Heritrix crawler. It can support broad crawls and is easily configurable, fast, tried and tested by several well-known cultural heritage institutions in the world. Heritrix was compared to other crawlers, namely Browsertrix and Brozzler. Browsertrix and Brozzler ensure high quality crawls but are slower and more experimental.
- The component providing access to captured Web archives is based on *WARCLight*. It consists of a catalog that is text searchable. Search criteria include year, public suffix, and content type. The replay search function is based on PyWB and allows for searching via URL or timestamp.
- The quality analysis, to ensure a semi-automatic control of the visual correspondence (does the captured website look the same as the source?), the interactional correspondence (can you interact the same way?) and the completeness (do we have every resource of the original?) is based on the structural similarity index (SSIM) and the visual quality indicator (VQI) algorithms and the comparison between successful requests in an archived website and requests in the original website. It balances the resource importance and takes factors into account such as content type, CSS coverage, image and size position.
- Derivative tools were needed to properly display the archives. The *Archives Unleashed Toolkit* was selected as it can produce graphs that represent the links between websites (GraphML). This tool can also explore the contents of Web documents as plain text.

Although the results of the project were positive and the necessary tools tested, they have not yet been incorporated into our digital archiving ecosystem. This requires significant resources, in personnel and other means, which are currently not available.

3.4 Externalization of some functions to federal institutional operators

As we cannot finance all the components of our digital archival system with our own resources, we promote shared solutions at the federal level with the support of the Belgian federal government. A central hub for such solutions is the „G-Cloud“. The idea is that one institution develops a service and can share it with other institutions through the G-Cloud platform.

Five years ago, the Belgian Scientific Policy Office (BELSPO) was convinced to build a shared long-term storage infrastructure for the 10 scientific and cultural heritage institutions under its authority. The State Archives participated in the requirement definition process. The resulting specifications were a compromise between all stakeholder expectations. The long-term preservation storage component of our archival system consists of an LTO tape library and is financially supported by BELSPO, technically maintained by a private consortium, but hosted in government data centers. This infrastructure will be included in the catalog of G-Cloud services.

We lobbied in favor of a shared federal records management platform, a solution for the preservation of semi-active digital records. Semi-active records storage is not part of the core business of the State Archives but the promotion of best practices in the public sector certainly is. Moreover, it is in our interest that records are well preserved until their transfer to our digital repository. We strongly supported the solution proposed by Smals, a non-profit organization that works for the Belgian public sector. Smals develops and integrates IT solutions. In this case, they combined integration of existing software and in-house developments to build the *Archiving as a Service* (Aaas) platform, which is based on the ARCSYS solution. Some functionalities should be added to the system, such as a connector with our digital archival infrastructure and a better generic access Web app for archive creators. Unfortunately, the connector with our system has been pending for three years because of budget problems.

Web archiving as a government mission will be attributed to the Royal Library. But regular consultation with the State Archives will ensure that the websites of all Belgian public sector institutions are archived according to the federal law on archives. The State Archives will oversee quality control for this category of websites, which will be accessible through our search engine as well.

4 Early, proactive intervention

Active and semi-active records storage is not the core business of a national archives, but to give advice on recordkeeping to civil servants fits within the archival supervision mission of the State Archives. We provide advice on content and supervise trainers for digital recordkeeping training programs, which are organized by the Federal Administra-

tion Training Institute. This program, called DIMA (*Digital Information Management Academy*), has a participatory approach, an on-the-job training methodology with workshops, and it offers a one-year coaching program during a records management project.

The Digital Act, which is the law of July 21, 2016, regulates the use of electronic trust services. It mainly concerns the electronic signature for citizens, the electronic seal for companies and other legal entities, the electronic timestamp, the electronic registered mail, and electronic archiving. This law is the Belgian translation of the European eIDAS regulation (EU 910/2014 *Regulation on electronic identification and trust services for electronic transactions in the European Single Market*). Except for electronic archiving, the Digital Act applies throughout the European Union.

Two important aspects of the Digital Act have a direct and indirect impact on the public sector and on the State Archives:

1. Digitization of paper documents. The law creates a new framework whereby digitized documents are given the same legal value as the paper originals. To this end, administrations will have to set up a digitization procedure that will be evaluated in advance according to the requirements of the law. The law distinguishes between a „qualified” and an „unqualified” electronic archiving service. Qualification requires certification. For documents that have been digitized by public administrations in recent years, transitional measures will have to be provided as the “qualified” certificate cannot be granted retroactively.
2. Digital retention of digital documents (both digital-born and digitized). Archive creators who are subject to the archive legislation must keep some of their records and transfer them to the State Archives after expiration of the retention period. In accordance with the principles of the new law, public authorities will have to resort to a qualified electronic archiving system for the preservation of these records. The law also references various international standards.

5 SWOT-Analyses

Despite small budgets, this approach has some advantages:

- The tools we get are tailor-made and fit our users’ expectations perfectly, even if technical or time constraints sometimes limit the quality and user-friendliness of the products.
- We can build a robust and valuable internal know-how to reduce our dependence on external factors and operators, for instance on private companies’ commercial policies.
- The reactivity of our development team is high and fast in case of bugs or breakdown. We do not use a ticketing system, and our experts can meet their colleagues within a few minutes to evaluate their problems.

On the other hand, the development of a solution is slow due to the small size of our technical team. For instance, it took 3 years from the user expectations analysis up to the launching of SAM. This is probably slower than the software development with an external subcontractor. But this observation may be mitigated by the fact that we spare the time of a public procurement procedure (that is at least 5 to 7 months). It should also be noted that the recruitment of IT experts is extremely difficult and slow on the Belgian job market, especially in Flanders, as these are occupations with skill shortages.

The fact that we can fund development of certain tools and know-how through research projects makes it possible to quickly acquire expertise from external partners such as universities and to achieve collaboration between multiple research entities with the State Archives as the unifying factor. We can propose our collections and our needs as case studies for research projects. This requires some degree of creativity when we must adapt our offers and expectations to the frame and timing of projects. Close attention must also always be paid to the practical outcomes delivered by the partners, particularly university researchers, whose professional environment and career are not necessarily oriented towards production, productivity, and daily public service. There is also a risk to lose a great part of our know-how when a colleague leaves the institution. And it sometimes happens! It is highly challenging to keep IT experts and digital archivists in house. We try to reduce the impact of this risk factor through a policy of knowledge sharing and transfer. Each application, each adjustment of an application must be properly documented. Each colleague is responsible for a group of tasks, but for each task they are responsible for, there is an understudy; a backup who can immediately take their place and execute the task in case of emergency.

We are tied to the priorities and planning set out by partners. Let’s think for example about *Archiving as a Service* (AaaS), whose connector to the archival system of the State Archives has been delayed for two years. This has had an impact on the functional analysis for the future Digital Transfer module in SAM. The same is true for the LTP platform, which is a de facto backup solution because some of the essentials of an archiving system were considered not to be priorities.

6 Conclusions

The transition from an „analog“ to a digital archive environment is not an easy task because the requirements for an ecosystem for digital archiving are still evolving depending on legal provisions and the development of technology. The approach of the State Archives was largely determined by the institutional and political context, with few recourses. This has meant that while we have been able to develop some aspects, there is not yet an integrated solution. We pay due attention to fill the gaps, nevertheless this process is slow but can be scaled up significantly if the requested funds are granted. We may conclude without shame that the construction of our digital archiving infrastructure suffers a 10-to-15-year delay compared to the other national archives. But this grants us the opportunity to learn from their successful and unsuccessful experiences, and even to reuse some of their tools. This allows us to scale up quickly and avoid unnecessary costs. Being a scientific institution allows us to have research projects funded, which gives us the opportunity to experiment with new technologies without affecting the normal operation of the service. These project results are evaluated and if favorable retained for implementation. There is still a long way to go, but we will surely get there.

Education and Training in Digital Recordkeeping in France

Édouard Vasseur, Centre Jean Mabillon, École nationale des chartes – PSL, Paris, France

My interest in the education and training in digital recordkeeping is based on my professional experience. In that regard, I would like to share with you a little personal anecdote. Two years ago, I was the functional head of the interdepartmental digital recordkeeping program Vitam. During a visit to the National Archives staff, with whom we were working on the implementation of the software we were designing, fellow archivists asked for my opinion on their work. The question dealt with the recovery of digital records acquired since the 1980s at the National Archives. They were processing a register produced in the form of a database, for which annual exports had been made. This database was the successor to paper registers that had existed since the 19th century. It became obvious during our conversation that they did not make the link between this database and the physical registers that had preceded the digitisation of this record. The database was being treated as a discrete item, without seeing the continuity of the series. But that was not the question being asked of me. Their interest was to retrieve the whole content of the database to produce a research tool in a xml file, without considering the confidentiality of the data it contained (the database related to people convicted in court, with the reason for their conviction). This experience had a profound effect on me, especially at a time when I was starting to apply for a professorship at the École nationale des chartes.

Since the beginning of my career, almost twenty years ago, digital technology and digital records have been everywhere. I started my career as an administrator of a documentary database. As head of the records management and archives office of the Ministry of Culture, I was responsible for digitising archival processes, setting up an recordkeeping policy for digital records and auditing the recordkeeping policy of organisations where digital technology was becoming increasingly important. As functional officer of the interdepartmental digital recordkeeping program Vitam, working daily with computer specialists, I had to define the functionalities of a software, test it, and present it to colleagues. And yet, my training in digital technology and digital recordkeeping was minimal: barely twenty hours of courses over five years of study.

I became more and more interested in how records managers and archivists were educated and trained today in the use of digital technologies in general and the recordkeeping of digital records in particular. This is the purpose of my paper.

This paper is based on an analysis of the professional literature and more particularly of the articles relating to education and training of records managers and archivists as well as the challenges awaiting them. Furthermore, this contribution deals with the courses offered by education and training organisations in light of this analysis.

I will first give a brief presentation of the training courses for records managers and archivists in France. Then, I will present some of these courses and evaluate them with regard to the recommendations made by the professional literature. Finally, I will report on the experiences of the last two years in the context of my teaching activities.

1 Education and training of records managers and archivists in France

Until the 1970s, the École nationale des chartes, where I am now a professor, was the only training body for archivists in France. The situation has now changed, and France has several bodies responsible for the education and training of records managers and archivists.

1.1 Education

Records managers and archivists in France are mainly educated in two organisations:

- the Institut national du patrimoine (INP)
- universities proposing master's degrees in archival science

The Institut national du patrimoine is an organisation created in 1990 and responsible for educating future curators hired by the Ministry of Culture and by regional and local authorities, who are destined to work in archives, museums or archaeological services¹. They are recruited through a competitive examination from people with a bachelor's degree – for archivists, mainly after a four-year course at the École nationale des chartes – and follow an 18-month training course, alternating between theoretical courses and internships. The theoretical courses include modules common to all future curators (law, finance, human resources, communication, restoration) and modules specific to each speciality (archives, museums, archaeology). On completion of the course, archive curators can work in state, departmental and municipal archive services, as well as in ministries.

1 Online at: <https://www.inp.fr/en> (retrieved on: 14th July 2023).

In addition, several universities offer master's degrees that educate records managers and archivists. There are currently nine master's degrees today, distributed as follows over the territory:

- Mulhouse, the oldest degree, founded in 1976²
- Lyon³
- Angers, founded in 1994⁴
- Toulouse⁵
- Versailles-Saint-Quentin en Yvelines⁶
- Aix-Marseille⁷
- Amiens, University of Picardy⁸
- Lille⁹
- Paris VIII¹⁰

These master's degrees are mainly attached to faculties of humanities, literature or history. These two-year courses are open to students with a bachelor's degree, who are recruited on the basis of their applications. The course combines theoretical and practical teaching and culminates, in the fourth semester, with an internship as a records manager or archivist, in a company or an archive service. The education can be dedicated to recordkeeping or be broader, and therefore can be relevant for other heritage or tourism professions.

There are also two more specialised master's degrees, which, although they do not strictly speaking educate records managers and archivists, prepare students who may be hired by archival services:

- The École nationale des chartes offers a master's degree in „Digital technologies applied to history“, which trains students in digital humanities¹¹
- The École nationale supérieure des sciences de l'information et des bibliothèques offers a one-year master's degree in digital recordkeeping and preservation, focusing on the preservation of digital objects and open to records management¹².

Finally, the Association de prévention du site de La Villette offers an educational program for people who have a bachelor's degree or are looking for a job and wish to reorient themselves in recordkeeping¹³.

1.2 Training

Training courses dedicated to records managers and archivists have been developed since the 1980s, mainly through the work of the Ministry of Culture and the Association of French Archivists. These courses do not provide certification.

The Association of French Archivists is undoubtedly the main player in this field, with a training centre run by a subsidiary, which offers around sixty courses for professionals each year, presented in a catalogue that is updated annually¹⁴.

The Ministry of Culture has for a long time offered about thirty courses, also intended for professionals, presented in a catalogue that is updated every year. Since 2020, the organisation of these training courses has gradually been entrusted to the École nationale des chartes and the Institut national du patrimoine.

2 Online at: <https://www.campus-fonderie.uha.fr/formations/histoire-patrimoine/master-parcours-gestion-contemporaine-des-archives/> (retrieved on: 14th July 2023).

3 Online at: <https://facdeslettres.univ-lyon3.fr/master-archives-2> (retrieved on: 14th July 2023).

4 Online at: <https://formations.univ-angers.fr/fr/offre-de-formation/master-lmd-MLMD/sciences-humaines-et-sociales-SHS/master-archives-IZFFWD2O.html> (retrieved on: 14th July 2023).

5 Online at: <https://ddame.univ-tlse2.fr/accueil/master-2-valari> (retrieved on: 14th July 2023).

6 Online at: <https://www.uvsq.fr/master-2-archives-parcours-gestion-des-archives-et-de-larchivage-gda> (retrieved on: 14th July 2023).

7 Online at: <https://allsh.univ-amu.fr/fr/formation/masters/master-histoire-civilisations-patrimoine/parcours-metiers-archives-bibliotheques-mediation-de> (retrieved on: 14th July 2023).

8 Online at: <https://www.u-picardie.fr/catalogue-formations/FI/co/master-hcp-mata.html> (retrieved on: 14th July 2023).

9 Online at: <https://www.univ-lille.fr/formations/fr-00002342.html> (retrieved on: 14th July 2023).

10 Online at: <https://www.univ-paris8.fr/-Master-Archives-628-> (retrieved on: 14th July 2023).

11 Online at: <http://www.chartes.psl.eu/en/cursus/digital-technologies-applied-to-history-master-s-degree> (retrieved on: 14th July 2023).

12 Online at: <https://www.enssib.fr/l-offre-de-formation/masters/archives-numeriques-presentation> (retrieved on: 14th July 2023).

13 Online at: <http://apsv.fr/formation/#assistant-e-archiviste> (retrieved on: 14th July 2023).

14 Online at: <https://www.archivistes.org/-Formations-> (retrieved on: 14th July 2023).

Two other organisations are also involved in the continuing education of records managers and archivists:

- the Conservatoire national des arts et métiers, which offers a specialised certificate in “Records management and digitalisation”¹⁵
- the Centre national de la fonction publique territoriale (CNFPT), which offers training for archivists in the regional and local authorities, particularly to prepare them for administrative competitions¹⁶.

Finally, mention should be made of the training courses available online:

- the courses offered by the Portail international archivistique francophone¹⁷
- those offered by the Club des responsables de politiques et de projets d’archivage with a MOOC on records management¹⁸ and a MOOC on the archiving of electronic messages¹⁹.

As you can see, France offers many training courses in this field.

2 Assessment criteria

What criteria should be used to evaluate education and training programs provided to students and archivists on digital recordkeeping and the use of digital technologies? This is the question I had to deal with in preparing this paper, and the answers are not yet fully satisfactory.

2.1 Existing tools

The professional literature on education and training of records managers and archivists has been limited in recent years, at least in French. The last issue of the French professional journal, *La Gazette des archives*, dedicated to educational issues dates back to 2010²⁰ and the last article devoted to the subject goes back to 2012²¹. It is in the proceedings of the University of Louvain’s meetings that we should look for some discussions on the subject in the French-speaking world. And yet these points are only outlined in the 2011 session devoted to digital projects²², in the 2014 session devoted to archives fifteen years later²³ and in the session devoted to the future archivist of 2030²⁴.

It is in the Anglo-Saxon world that we find the most important discussions, in particular those carried out within the framework of the InterPARES program and its latest version IPTrust, as illustrated by the latest book published under the direction of Luciana Duranti and Corinne Rodgers with an entire article devoted to the subject of education and training²⁵.

In addition to the professional literature, I therefore consulted the professional frame of reference published by the Association of French Archivists, the second version of which was released in 2017²⁶. It would also have been useful to have a closer look at the job descriptions offered to young professionals and current archivists, but this work could not be carried out in the context of preparing this paper.

Consultation of the professional literature and the professional frame of reference of the Association of French Archivists does, however, make it possible to identify the knowledge, skills and attitudes that seem necessary for records managers and archivists in today’s world.

15 Online at: https://formation.cnam.fr/rechercher-par-discipline/certificat-de-specialisation-records-management-et-dematerialisation-200631.kjsp?RF=newcat_themes (retrieved on: 14th July 2023).

16 Online at: <https://www.cnfpt.fr/> (retrieved on: 14th July 2023).

17 Online at: <https://www.piaf-archives.org/tous-les-cours> (retrieved on: 14th July 2023).

18 Online at: <http://blog.cr2pa.fr/mooc/> (retrieved on: 14th July 2023).

19 Online at: <http://blog.cr2pa.fr/2017/06/mooc-sur-les-mails-95-videos-gratuitement-a-votre-disposition/> (retrieved on: 14th July 2023).

20 Online at: https://www.persee.fr/issue/gazar_0016-5522_2010_num_218_2 (retrieved on: 14th July 2023).

21 Anne-Marie Bruleaux, Former l’archiviste du XXIe siècle, in: *La Gazette des archives*, 226/2 (2012). Gérard Naud, un archiviste de notre temps, Marie-Edith Enderlé-Naud (Ed.), pp. 71-93, (online at: https://www.persee.fr/doc/gazar_0016-5522_2012_num_226_2_4898 (retrieved on: 14th July 2023).

22 Nicolas Delpierre, François Hiraux, Françoise Mirguet (eds.), *Les chantiers du numérique, dématérialisation des archives et métiers de l’archiviste*, Louvain-la-Neuve 2012.

23 Paul Servais, Françoise Mirguet (eds.), *L’archive dans quinze ans: vers de nouveaux fondements [actes des quatorzièmes Journées des archives, Louvain-la-Neuve, 24-25 avril 2014 / organisée par le service des Archives de l’Université catholique de Louvain]*, Louvain-la-Neuve 2015.

24 Paul Servais, Françoise Mirguet (eds.), *Archivistes de 2030: réflexions prospectives*, Louvain-la-Neuve 2014.

25 Luciana Duranti, Corinne Rogers (eds.), *Trusting records in the cloud*, Londres 2019.

26 Online at: <https://www.archivistes.org/Referentiel-metiers-et-fiches-de-poste> (retrieved on: 14th July 2023).

2.2 Cross-cutting skills

I will first mention the case of cross-cutting knowledge and skills, specifically those that are not necessarily dedicated to digital recordkeeping and the use of digital technologies, but which are considered essential for today's records manager and archivist.

The following skills may be grouped in this category

- the ability to define the scope of the objects to be captured in a recordkeeping system using the main principles of diplomatics, while taking into account the challenges of preservation and access;
- knowledge of the main concepts and functions of recordkeeping
- the ability to analyse and model a business process;
- an introduction to project management, risk management, audit methodology, managing relationships with a service provider;
- the capacity to express a need that can be translated into the construction of a building or a tool, including digital ones.

2.3 Specific skills

The following knowledge can be grouped in this category:

- a minimum knowledge of digital technologies, of the ecosystem of the digital world (what does a Chief Technical Officer do? A Chief Information Security Officer? A Data Analyst?) and the way documents and records are created, exchanged and stored today
- an understanding of at least the language of computer scientists
- a minimum knowledge of IT law, particularly in terms of personal data protection and authenticity
- a minimum knowledge of IT security rules, in particular user rights management
- an understanding of the role of metadata
- familiarity with the challenges of preserving digital objects
- knowledge the techniques of dissemination on the Internet
- an understanding of the role of social networks and how to use them
- familiarity with current IT topics: cloud computing, software as a service, blockchain, data lake.

3 Presentation of a selection of education and training programs

Let us now quickly analyse a few training programs. As the data I have collected is not exhaustive, I will limit myself to giving a few examples.

3.1 The example of archive curators trained at the Institut national du patrimoine (INP)

The curators recruited through a competitive examination are trained for eighteen months, including nine months of internship.

The students receive, on the one hand, the same core courses as all the students and, on the other hand, courses specific to their speciality (archives, museums, historical monuments, archaeology).

The common courses are as follows:

- ethics
- law
- economics of heritage
- finance
- human resources
- conservation and restoration
- building construction
- dissemination and promotion of heritage.

The courses specific to archivists (110 hours) are as follows:

- institutional meetings
- scientific and technical control, appraisal and acquisition
- arrangement and description
- digital recordkeeping
- conservation
- access and dissemination.

Among these courses, those more specifically devoted to digital recordkeeping and digital technologies are the following:

- a module on arrangement and description: practical use of software (4 hours)
- a seminar dedicated to digital recordkeeping, from its production to its preservation: development of e-government, standards and actors, role of metadata, acquisition and processing of data in an digital recordkeeping system, softwares, perpetuation (19 hours)
- a module on access: digitisation, website, web dissemination strategy (8 hours).

3.2 The example of master's degrees

I have chosen two master's degrees in archival science as examples, one in Angers and one at the University of Picardy in Amiens.

The master's degree in archival science at the University of Angers lasts two years²⁷. The courses relating to digital recordkeeping and digital technologies are spread over the two years of teaching:

- In the first year (228 hours plus 113 hours), students are introduced to archival science, diplomatics, palaeography, institutional history and research. A whole teaching unit is dedicated to digital technologies applied to the field of recordkeeping (analysis of the digital document, theoretical and practical presentation of software used in archival services such as AtoM), a second unit to digital technologies applied to the field of research (general presentation of tools used to acquire, analyse and visualise information, including in the form of maps and databases) and a third unit to computer tools likely to help students write their research dissertation (office automation) and publish texts (TEI, EAD).
- In the second year (348 hours plus internships), students broaden their knowledge of institutional history, diplomatics and archival science and can choose a specialisation in records management, access and dissemination, or archives management, within the framework of a project. An education unit is however dedicated to digital information and includes both courses dedicated to digital recordkeeping and to the dissemination of digital objects in the broad sense. Project management is present in several ways: students set up a study day, carry out a business project (records management, access and dissemination, preservation) in which digital technologies are systematically present. On the other hand, internships are not necessarily dedicated to digital recordkeeping projects or projects involving digital technologies.

The master's degree at the University of Picardie, less specialised in archival science but more open to historical research, also lasts two years²⁸. The courses relating to digital recordkeeping and digital technologies are spread over the two years of the program:

- In the first year, students are introduced to the history of institutions and archives, to archival science, to historical research and to the various heritage professions. In this context, they attend a project management workshop and are introduced to digital technologies (databases and digital cartography).
- In the second year, students are trained in records management and the use of digital technologies by the archivist (108 hours out of 194 hours). No teaching unit is dedicated to digital archiving.

3.3 An example of training

When I look at training courses, and in particular the catalogue of the Association of French Archivists, the situation is significantly different. Of the sixty or so courses offered, almost a third are dedicated to digital archiving and digital technologies, with courses that are easier to identify²⁹.

Among these courses, it is worth mentioning:

- business process analysis, identification of objects to be captured in a recordkeeping system
- basic knowledge of digital technologies: introduction to digital recordkeeping
- IT law: protection of personal data
- project management using digital technologies: design of an ECM, design of an electronic records management system
- preservation of digital objects
- techniques for enhancing archives on the internet: open data, digitisation, use of Wikipedia
- internet communication techniques: website design, use of social networks.

27 Online at: <https://formations.univ-angers.fr/fr/offre-de-formation/master-lmd-MLMD/sciences-humaines-et-sociales-SHS/master-archives-IZFFWD2O.html> (retrieved on: 14th July 2023).

28 Online at: <https://www.u-picardie.fr/catalogue-formations/FI/co/master-hcp-mata.html> (retrieved on: 14th July 2023).

29 Online at: <https://www.archivistes.org/-Formations> (retrieved on: 14th July 2023).

4 Evaluation of existing programs

What conclusions can be drawn from the analysis of the initial and ongoing training of records managers and archivists in France? Here I combine the results of my observations with the feedback from interviews conducted with several of those responsible for these education and training programs.

4.1 General conclusions

From a general point of view, there is no doubt that education and training organisations, especially education institutions, are striving to consider digital recordkeeping and the use of digital technologies in their courses. However, they are subject to significant constraints. On one hand, the number of hours to be devoted to theoretical teaching remains limited, especially when the training is not strictly related to recordkeeping. On the other hand, not all students are necessarily familiar with digital technologies or may tend to restrict them to the use of social networks. However, educational managers note that in recent years, teaching electronic recordkeeping and the use of digital technologies has been facilitated by the fact that employers are making that knowledge a priority in their recruitment.

First observation: it is logically easier for training organisations to offer courses dedicated to digital recordkeeping and digital technologies than it is for educational institutions.

Second observation: educational managers do not necessarily choose to offer courses dedicated to digital recordkeeping. Specific courses exist in Angers, Mulhouse or at the Institut national du patrimoine, while in Amiens, it was decided to reaffirm that digital recordkeeping is above all recordkeeping and that it should therefore be considered in a global way. In Angers, the offer of courses dedicated to digital recordkeeping is deliberately chosen, in order to clearly indicate to students who wish to apply for the master's degree that digital technologies will be an essential part of their courses. This solution makes it possible to exclude students who are not interested in digital technologies. All the educational managers believe that it is necessary to first provide a theoretical professional knowledge base before presenting the specificities of digital recordkeeping.

Third observation: the master's degrees more specifically dedicated to digital technologies (École nationale des chartes and École nationale supérieure des sciences de l'information et des bibliothèques) have a totally different approach. The former one focuses very clearly on digital technologies rather than on digital recordkeeping, which has, in reality, a very limited place in the curriculum. The latter has a focus on management and preservation of digital objects, without providing fully comprehensive education in recordkeeping.

4.2 Major Trends

A more detailed analysis then makes it possible to make three observations that are more precise on the content of the courses given.

Firstly, education places more emphasis on description, access and dissemination of archives using digital technologies rather than on records management. Digitisation and online publication of digital archives, publication of finding aids in xml-ead format have long been the main focus of training in digital recordkeeping and the use of digital technologies in archives. It is clear to educational managers that this is the most practical approach for students, as it makes it easier to capture their attention. This tendency is naturally reinforced when the educational manager is not an archivist but a historian. This emphasis on access and dissemination topics is also reflected in the orientation of the IT law course and in the presentation of privacy issues.

Secondly, education now systematically includes courses on project management. This teaching is implemented into practice in several ways, either in the form of dedicated teaching units (Toulouse), or via the organisation by students of a study day as part of their academic year. It should also be noted that some master's degrees, such as that of Versailles-Saint-Quentin, now emphasise other cross-cutting skills such as auditing or risk management. Educational managers insist that the introduction of project management into the curricula is more the result of political expectations formulated by the Ministry of Higher Education than of the expectations of professional archivists.

4.3 Weaknesses

French archivists are therefore trained in digital recordkeeping and digital technologies. Analysis of training courses – even if this would require further study – and experience show, however, that certain aspects of the subject matter are still insufficiently covered and do not adequately prepare future records managers and archivists for the evolution of their profession. This undoubtedly explains the number of courses dedicated to digital recordkeeping and digital technologies in training programs.

Main issue: the training given on the production of digital documents and records themselves and on records management remains insufficient. Analysis of digital documents and digital diplomatics remains limited. Students are not very

familiar with the digital ecosystem (knowledge of the various jobs of CTO, CISO, DPO, and so on) and with the current challenges of information technology (cloud computing, artificial intelligence, blockchain and so on). Process modelling is not very visible in the programs and in the skills put forward by the master's degrees, even if, according to some educational managers, it is present in the form of exercises and case studies. Training in IT law in general and on the General Data Protection Regulation (GDPR) in particular is fairly revealing of this observation: the courses place more emphasis on issues related to access and dissemination than on those related to management in organisations, even if some master's courses invite Data Protection Officers (DPOs). The place of records management in training courses has nevertheless increased in recent years, particularly in Angers and Saint-Quentin. However, it remains insufficient in my opinion.

Second issue: digital technologies sometimes seem to be approached more from the point of view of standards and tools than from that of needs. This situation is undoubtedly not widespread (cf. Toulouse), but it seems to be relatively present and to meet the expectations of the students themselves and even of the employers, despite a more nuanced discourse from the educational managers.

Third issue: although project management is now well represented in education, risk management and approaches that associate students more closely with IT professionals are still very little developed. For educational managers, the very limited number of training hours, administrative constraints, and the compartmentalisation of universities – not all universities with a master's degree in archival science have a master's degree in computer science – prevent the development of this approach. The involvement of computer specialists in courses and the emphasis placed on digital archiving and digital technologies in the context of internships are, however, a first means of encouraging a professional rapprochement between archival science and computer science.

5 Experiments conducted over the past two years

I would like now mention experiments conducted over the last two years at the École nationale des chartes and at the master's degree of the university of Versailles-Saint-Quentin-en-Yvelines.

At the École nationale des chartes, we do not educate students in recordkeeping, strictly speaking. Since 1990 that has been the mission of the Institut national du patrimoine. Nevertheless, as some students of the École nationale des chartes are destined to participate in the competitive examination of the Institut national du patrimoine, we need to give them an introduction to archival studies.

I am responsible for a 60-hour course in the history of institutions, diplomatics and archival studies, spread over three years. The aim is to make my students aware of the ways records and archives have been transformed in recent years, but at the same time to emphasise that good recordkeeping policies must embrace both traditional and digital records and archives.

That is why diplomatics is so important in my courses. The first year, students study diplomatics of 19th or 20th century. In the third year, three courses are delivered: firstly, the way documents and records have been created and transmitted since the beginning of the 19th century, including the development of computers and digital technology; secondly, digital diplomatics; thirdly, a case study with a specific type of record, in order to study how technical and organizational changes have affected the way records have been produced and transmitted since the 19th century (in 2020, diplomatic correspondence, in 2021, the landregister). My aim with these courses is to provide students with the tools to place the development of digital technologies in a long-term perspective, and to help them analyse digital documents and records. Experience shows, however, that students are still intimidated by digital records and that they struggle to make the connection between records they produce themselves every day, digital humanities and the everyday work of a records manager or an archivist.

The master's degree in archival science at the University of Versailles-Saint-Quentin-en-Yvelines is also an opportunity to experiment with new approaches. At this institution, I am responsible for two courses in diplomatics, one of them dedicated to digital diplomatics, as well as for one course on audit methodology. This year, with my colleague – head of the archives départementales of Yvelines – we decided to experiment with an integrated exercise in digital diplomatics and recordkeeping. The students had to analyse the process of allocating the active solidarity income (revenu de solidarité active). They had to gather the available documentation on this allocation, model the process, analyse the form taken by its documents and records, use their knowledge within diplomatics, and define the retention period and disposition of these documents and records. The first results are encouraging but show that students still have difficulties taking into account data stored in databases in their appraisal. These are points that need to be improved next year, but the exercise has the merit of inviting the students to consider in a global way the recordkeeping process and to include everything, data, documents and records.

6 Conclusion

What conclusions can be drawn from this analysis? That digital recordkeeping and digital technologies are well present in the training and education of records managers and archivists in France which allows today the implementation of ambitious electronic archiving programs. That these programs either have a clear focus on digital recordkeeping and digital technologies giving candidates a clear understanding of what awaits them in their professional lives or places greater emphasis on recordkeeping, regardless of the medium of the records. That the anchoring of archivist training in the faculties of humanities and social sciences tends to lock the training of archivists into a heritage vision of archives and does not place enough emphasis on records management.

We are clearly in the „neither too much nor too little“ situation feared by Diane Baillargeon a few years ago in Louvain³⁰, but the risks pointed out by Marie-Anne Chabin³¹ – having records managers and archivists focusing either on a physical vision of digital records or converted to computers and IT development and rejecting their competence in the document – cannot yet be ruled out.

There is no doubt that we should give more attention to education and training issues, particularly to prevent digital recordkeeping from turning mostly to an IT topic. Two points seem essential to me: to insist on the continuity between physical and digital records – in short, to insist that records are above all the traces of interactions between stakeholders – and to make future records managers and archivists highly aware of digital technologies (without being IT professionals). It is under these two conditions that we will avoid questions such as those raised by colleagues from the National Archives in the future.

30 Diane Baillargeon, De quelle sorte d'archivistes aurons-nous besoin en 2030?, in: *Archivistes de 2030* (see ann. 24), p. 20–32.

31 Marie-Anne Chabin, L'archiviste de 2030 entre archives numériques et utilisateurs connectés, in: *Archivistes de 2030* (see ann. 24), p. 61–72.

Die Ausbildung in (elektronischer) Schriftgutverwaltung und digitaler Langzeitarchivierung in Deutschland

Sarah Rudolf, Gregor Patt, LVR-Archivberatungs- und Fortbildungszentrum, Pulheim

1 Einleitung: Aus- und Fortbildungsstätten in Deutschland

In Zeiten der fortschreitenden Digitalisierung, in Zeiten einer weltweit verheerenden viralen Pandemie, in denen z. B. in Deutschland dennoch Fördermittel für die Kultur und konkret für die Digitalisierung von Kulturerbe in Archiven, Bibliotheken, Museen, weiteren Gedächtniseinrichtungen vom Bund zur Verfügung gestellt werden, in Zeiten des E-Government¹; in solchen Zeiten kann es nicht überraschen, dass auch die Fragen elektronischer Schriftgutverwaltung oder des elektronischen Records Management und der digitalen Archivierung trotz jahrzehntelanger Geschichte immer noch mehr an Bedeutung gewinnen.² Je mehr Daten entstehen, desto größer ist die Verantwortung der Wissenschaftsbetriebe, Kulturbetriebe und Gedächtniseinrichtungen, das Erhaltenswerte auch verfügbar zu halten.³

Wie Édouard Vasseur in seinem Beitrag zur Situation in Frankreich aufzeigt, haben Records Management und digitale Archivierung wie zu erwarten sowohl in Deutschland als auch in Frankreich in den letzten 30 Jahren kontinuierlich an Bedeutung gewonnen. Beide Themenkomplexe sind heute längst fester Bestandteil des archivarisches Alltags. Und dennoch sind nicht allorts gleichermaßen Kenntnisse und Kompetenzen vorhanden. Ältere Sorgen, wie die von Peter M. Toebak 2002 veröffentlichten, finden wir – beinahe glücklicherweise – zwar nicht überall in dieser Form bestätigt:

Vorauszusehen ist, dass E-Government als Konzept wiederum von den falschen Leuten vereinnahmt und in die falsche Richtung gelenkt wird. Einzige Hoffnung wäre, dass sich die I+D-Berufswelt (Archivare, Records Manager, Dokumentare und Bibliothekare) selbstbewusst steuernd in diesen Prozess einschaltet [...].⁴

Einen wahren Kern haben sie aber doch. Tatsächlich gehen die Kenntnisse in rechtssicherer Schriftgutverwaltung eher zurück. In einem föderalen Staat wie der Bundesrepublik ist dies eine Herausforderung, die Kommunen, Ländern und Bund gleichermaßen betrifft. Alle Akteur/innen haben deshalb in den vergangenen Jahren verschiedene „Auffangprogramme“ für diejenigen ins Leben gerufen, die ihre Ausbildung bereits abgeschlossen haben und für die es gilt, etwa aufgrund rasanter Innovationsprozesse, Wissenslücken zu schließen.⁵ Was die Ausbildung angeht, so liegt die Verant-

- ¹ Zu den einzelnen Fördermaßnahmen im Rahmen des Gesamtprogramms Neustart Kultur, online unter <https://neustartkultur.de/> (abgerufen am: 14.07.2023). Für den vertiefenden Einblick: die Seiten des Bundes zum E-Government, online unter: https://www.verwaltung-innovativ.de/DE/Startseite/startseite_node.html (abgerufen am: 14.07.2023), die UN-Resolution Schutz Kulturerbe, online unter: [https://www.undocs.org/S/RES/2347%20\(2017\)](https://www.undocs.org/S/RES/2347%20(2017)) (abgerufen am: 14.07.2023) und die UNESCO/ICA-Erklärung zu Archiven, online unter: https://www.ica.org/sites/default/files/UDA_Sept%202013_press_GE.pdf (abgerufen am: 14.07.2023); sowie einführende Literatur, hier bes. Frank M. Bischoff, Elektronische Registratursysteme, Überlegungen zur Archivierung digitaler Unterlagen der nordrhein-westfälischen Verwaltungsgerichte [AUDS-Tagung] 1997; online unter: https://www.staatsarchiv.sg.ch/home/auds/01/_jcr_content/Par/downloadlist_0/DownloadListPar/download_2.ocFile/Text%20Bischoff.pdf (abgerufen am: 14.07.2023).
- ² Peter M. Toebak, Wissensmanagement und Dokumentenmanagement, Überlegungen aus der Praxis, in: *Arbido 2* (2002), S. 18–20; Michael Wettengel, Verwaltungsmodernisierung und IT-gestützte Vorgangsbearbeitung in der Bundesverwaltung, online unter: <https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2000/it-entwicklungen-/21-Text%20Wettengel%20Verwaltungsmodernisierung.pdf> (abgerufen am: 14.07.2023); Nils Brübach, Manuela Queitsch, Hans Liegmann †, Achim Oßwald, 4.2 Das Referenzmodell OAIS, in: H. Neuroth / A. Oßwald / R. Scheffel / S. Strathmann / K. Huth (Hrsg.): *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3.* Göttingen 2010; online unter: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf (abgerufen am: 14.07.2023); Sara Marty, Archive und Bibliotheken für das Kulturerbe, in: *arbido 1* (2018), online unter: <https://arbido.ch/de/ausgaben-artikel/2018/archive-und-bibliotheken-f%C3%BCr-das-kulturerbe/editorial-archiv-e-und-bibliotheken-f%C3%BCr-das-kulturerbe> (abgerufen am: 14.07.2023); Gesa Bierwerth, Das kulturelle Erbe, ein junges Konzept, in: *arbido 1* (2018), online unter: <https://arbido.ch/de/ausgaben-artikel/2018/archive-und-bibliotheken-f%C3%BCr-das-kulturerbe/das-kulturelle-erbe-ein-junges-konzept> (abgerufen am: 14.07.2023).
- ³ Patrick Sahle, Jonathan Blumtritt, Forschungsdatenmanagement als Puzzlespiel? Institutionelle Aufgaben und Rollen bei der Versorgung der Geistes- und Kulturwissenschaften, in: *Archivar 2020* (73) 1, S. 19–24. In diese Richtung weisen auch das institutionelle Engagement und die Positions- sowie Empfehlungspapiere des Rats für Informationsinfrastrukturen, online unter: <https://rfii.de/de/dokumente/> (abgerufen am: 14.07.2023): Zu nennen ist v. a. das Empfehlungspapier „Digitale Kompetenzen – Dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft“.
- ⁴ Toebak (wie Anm. 2), S. 20.
- ⁵ Als ein Beispiel mit dem regionalen Bezug zu NRW sind die sog. E-Government-Grundlagenprojekte (EGP) der Landesregierung zu nennen, mit ihren verschiedenen Arbeitspaketen, an denen neben verschiedenen Ministerien und Behörden der Landesverwaltung teilweise auch das Landesarchiv NRW mitwirkt. Hierfür und auch für Einstellung neuer Fachkräfte wurden vom Land ab 2018/19 umfassende Mittel zur Verfügung gestellt. Vgl. hierzu die Informationen auf den Seiten des NRW-Chief Information Officer (CIO), online unter: <https://www.e-verwaltungsarbeit.nrw> (abgerufen am: 14.07.2023).

wortung für die allgemeine verwaltungsfachliche Ausbildung bei den jeweiligen Ausbildungsbehörden in Verbindung mit den Hochschulen für öffentliche Verwaltung, soweit es z. B. den allgemeinen staatlichen und kommunalen Verwaltungsdienst betrifft. Sobald es um die Laufbahnen von Archivar/innen im gD und hD geht, sind es die staatlichen Archive des Bundes und der Länder, die in Verbindung mit der Archivschule Marburg, Hochschule für Archivwissenschaft, bzw. in Bayern mit der Bayerischen Archivschule in München ausbilden. Hinzu kommen Hochschulen und Universitäten, die Studiengänge z. B. zur Public Administration und zu der Archivwissenschaft nahestehenden Fächern wie der Bibliothekswissenschaft anbieten. Auch Berufs- und Fachverbände tragen dazu bei, Berufsethik, Anforderungen, Standards zu formulieren und zu etablieren.⁶ Archive aller Sparten agieren zudem oft im Rahmen der sogenannten „Behördenberatung“ als Kompetenzzentren für Fragen der Schriftgutverwaltung für ihre Träger. Eine besondere Funktion kommt in Nordrhein-Westfalen den Archivämtern der beiden Landschaftsverbände zu (LVR-Archivberatungs- und Fortbildungszentrum (LVR-AFZ), LWL-Archivamt für Westfalen). Sie beraten mehrere Hundert Archive unentgeltlich bei allen Fragen rund um die Archivierung ihrer Bestände und den Betrieb von Archiven mit dem Ziel, ihnen eine möglichst angemessene sachliche, personelle und finanzielle Grundausstattung zu verschaffen. Beratung zu Records Management und digitaler Archivierung sind integraler Bestandteil dieser Dienstleistungen für die nordrhein-westfälischen Archive.⁷

Der Fokus der folgenden Darstellung liegt auf den Angeboten der überregionalen Akteure Deutsche Universität für Verwaltungswissenschaften in Speyer⁸, Archivschule Marburg⁹ und Fachhochschule Potsdam mit zwei Fernstudiengängen zur Archivwissenschaft¹⁰. Ausgehend von der Situation in NRW werden zudem exemplarisch ein Studiengang der Bergischen Universität Wuppertal sowie die Angebote der Hochschule für Polizei und öffentliche Verwaltung in NRW¹¹ vorgestellt. Außerdem soll ein kurzer Blick auf verschiedene Fortbildungsangebote geworfen werden.

2 Schriftgutverwaltung, digitale Archivierung: Ausbildungsumfang und -inhalte

2.1 Universitäten, Verwaltungs(fach)hochschulen, Fortbildungsangebote

Lehre und Forschung unterliegen in Deutschland keinerlei Einschränkungen. Während der Fächerkanon mit seinen Inhalten also an Universitäten und Fachhochschulen zu großen Teilen in Eigenverantwortung verwaltet, angepasst und verändert werden kann – ministerielle Genehmigungen vorausgesetzt –, so unterliegt die Ausbildung in Verwaltungen, an Verwaltungsfachhochschulen und spezialisierten Fachhochschulen wie der Archivschule Marburg komplett in der Hand der für die Ausbildung verantwortlichen staatlichen oder kommunalen Behörden.¹²

Verschiedene Verwaltungsreformen im Zuge des New Public Management und anderer ausbildungsbezogener Reformprozesse haben dazu geführt, dass die Studienanteile zur Schriftgutverwaltung in der Ausbildung der Verwal-

6 Jean Dryden, Archival Associations as Standards-Setting Bodies, in: Journal of Archival Organization 4 (2006) 3/4, S. 133–138; Matthew R. Francis, 2013 Archival Program Graduates and the Entry-Level Job Market, in: The American Archivist 78 (2015) 2, S. 514–547; Didier Grange, Les associations professionnelles d'archivistes. Panorama international. Paris 2016 (= La Gazette des Archives 241 [2016] Nr. 1); Robert Kretzschmar, Ein Fachverband – ein Berufsbild? Zum Selbstverständnis des Verbands deutscher Archivarinnen und Archivare, in: Karsten Uhde (Hrsg.), Berufsbild im Wandel – Aktuelle Herausforderungen für die archivarische Ausbildung und Fortbildung. Beiträge zum 9. Archivwissenschaftlichen Kolloquium der Archivschule Marburg. Marburg 2005, S. 126–199; Ders., Profil und Professionalisierung eines archivarischen Berufs- und Fachverbands. Der VdA – Verband deutscher Archivarinnen und Archivare e. V., in: Marcus Stumpf (Hrsg.), Beruf und Berufsbild des Archivars im Wandel. Münster 2008, S. 45–56; Caroline Sampson, Beruflich-fachliche Normierung, in: 50 Jahre Verein deutscher Archivare. Bilanz und Perspektiven des Archivwesens in Deutschland. Siegburg 1996, S. 259–264.

7 Siehe hierzu die Informationsseiten der beiden Archivämter, online unter <https://afz.lvr.de/de/startseite.html> und <https://www.lwl-archivamt.de/de/> (abgerufen am: 14.07.2023); sowie die Landschaftsverbandsordnung: § 5 Abs. 1b Nr. 3 LVerbO, online unter: https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=2022&bes_id=4988&menu=0&sg=0&aufgehoben=N&keyword=landschaftsverbandsordnung#det0 (abgerufen am: 14.07.2023).

8 Online unter: <https://www.uni-speyer.de/> (abgerufen am: 14.07.2023).

9 Online unter: <https://www.archivschule.de/DE/ausbildung/> und <https://www.archivschule.de/DE/fortbildung/> (abgerufen am: 14.07.2023).

10 Online unter: <https://www.fh-potsdam.de/studium-weiterbildung/studiengaenge> (abgerufen am: 14.07.2023).

11 Online unter: <https://www.hspv.nrw.de/> (abgerufen am: 14.07.2023).

12 Siehe hierzu sowohl das Gesetz über die Hochschulen des Landes NRW (Hochschulgesetz – HG NRW) als auch die Verordnung zur Regelung des Näheren der Studienakkreditierung in Nordrhein-Westfalen (Studienakkreditierungsverordnung – StudakVO), online unter: https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=221&bes_id=28364&menu=0&sg=0&aufgehoben=N&keyword=hochschulgesetz#det0 und https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=00420200401070525504 (abgerufen am: 14.07.2023).

Zum Vergleich auch das Gesetz über die Fachhochschulen für den öffentlichen Dienst im Lande Nordrhein-Westfalen (Fachhochschulgesetz öffentlicher Dienst – FHGöD), online unter https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=221&bes_id=4445&menu=0&sg=0&aufgehoben=N&keyword=hochschulgesetz#det442155 (abgerufen am: 14.07.2023); sowie die Rechtsgrundlagen der Archivschule Marburg, online unter <https://www.archivschule.de/DE/wir-ueber-uns/rechtsgrundlagen/> (abgerufen: am 14.07.2023).

tungskräfte weitestgehend entfallen und wissenschaftliche Grundlagen in den Vordergrund gerückt sind, während die Thematik parallel zu dieser Entwicklung im Berufsalltag und in der Ausbildung von Archivar/innen zunehmend ausgeweitet worden ist (siehe Kap. 2.2). Denn Fragen digitaler Archivierung, ja der Archivierung allgemein, werden in allgemeinen verwaltungsfachlichen Ausbildungen ebenfalls nicht explizit in den Modulen benannt und hängen damit zunehmend von der Ausgestaltung praktischen Studien ab.¹³

Das gilt ebenso für entsprechende Studiengänge der heute von Bund und Ländern gemeinsam getragenen Deutschen Universität für Verwaltungswissenschaften in Speyer. Wie auch an den Verwaltungshochschulen der Länder sind die Module schwerpunktmäßig der Lehre und Ausbildung im New Public Management gewidmet. Belange von Records Management und Archivierung sind nicht zu greifen. Exemplarisch sei auf den Master of Arts Public Administration (MAPA) der Deutschen Universität für Verwaltungswissenschaften in Speyer verwiesen. Die Module heißen 1) Sozialwissenschaft I, 2) Öffentliches Recht, 3) Wirtschaftswissenschaften, 4) Methoden der empirischen Sozialforschung I, 5) Perspektiven des interdisziplinären Arbeitens, 6) Sozialwissenschaft II, 7) Methoden der empirischen Sozialforschung II, 8) Information, Kommunikation, Handlungskompetenz, 9) Wirtschaftswissenschaften, 10) Regieren und Verwalten, 11) Europäisierung und Internationalisierung, 12) Management und Organisation und 13) Studium generale.¹⁴

Dass Records Management und digitale Archivierung auch in anderen universitären Fächern, die der Archivwissenschaft nahestehen (z. B. Geschichtswissenschaften, Germanistik, Rechtswissenschaften, historische Hilfswissenschaften), bestenfalls am Rande vorkommen, überrascht im Vergleich zu den Verwaltungswissenschaften eher nicht. Bemerkenswert ist vielmehr, dass jüngere Entwicklungen wie die Digital Humanities möglicherweise eine kleinere Trendwende eingeleitet haben. So werden in den Modulen des Masterstudiengangs Editions- und Dokumentwissenschaft (EDW) der Bergischen Universität Wuppertal Fragen nach archivarisches Kompetenzen sowie Aufgaben, und wie sie in Verbindung zur geisteswissenschaftlichen Arbeit stehen, explizit berücksichtigt.¹⁵ Die Wahlpflichtmodule I, II und III benennen die Themen zwar nicht namentlich, doch sind sie thematisch verhältnismäßig offen gefasst, sodass hier archivwissenschaftliche Themen platziert und über Kooperationen abgedeckt werden können.¹⁶ An dieser Universität ist das LVR-AFZ schon seit Jahren über eine Lehrbeteiligung damit befasst, Studierenden Kompetenzen, Tätigkeitsfelder und Aufgaben von Archiven zu vermitteln. Jüngst konnte die Verfasserin im Rahmen eines Lehrauftrags für den Wahlpflichtbereich im Master EDW Anforderungen der fachgerechten Digitalisierung, digitalen Archivierung und Präsentation von Digitalisaten, Editionen, Forschungsergebnissen vermitteln. Die Dokumentation der Projekte und damit implizit eine Form des möglichst transparenten Records Managements war ein wichtiger Bestandteil des Kurses.

Es ist auch zu beobachten, dass hoch spezialisierte Archivsparten wie z. B. Literaturarchive immer stärker den Bezug zur Wissenschaft suchen, um die jeweiligen fachlichen Anforderungen und Erwartungen über Kooperationen miteinander in Bezug setzen zu können. Das gilt nicht zuletzt für die Anforderungen und Erwartungen in der digitalen Archivierung, angefangen bei Fragen der Bewertung und Übernahme.

Weil sowohl Verwaltungshochschulen als auch wissenschaftliche Hochschulen und die Universitäten jeweils andere Schwerpunkte setzen, sind Kooperationen und Fortbildungsangebote als Weg der Vermittlung von fachlichen Standards und Wissen über Records Management und digitale Archivierung stärker in den Fokus geraten. Sowohl der Verband deutscher Archivarinnen und Archivare e. V. (VdA)¹⁷, das Kompetenznetzwerk „nestor“ als auch die Archivschule Marburg und die Archivämter der nordrhein-westfälischen Landschaftsverbände berücksichtigen diese Tatsache, indem sie unterschiedliche Arten von Fortbildungsveranstaltungen in ihre Angebote aufgenommen haben. So bietet die Archivschule Marburg seit Jahren ein jährliches Fortbildungsprogramm mit Grund-, Aufbau- und Erweiterungs-

13 Siehe hierzu die Modulübersicht für den kommunalen Verwaltungsdienst i.d.F. vom 14.07.2023, online unter: https://www.hspv.nrw.de/dateien_studium/studium-und-lehre/BA/kvd/moduluebersicht/50_Moduluebersicht_KVD_ab_EJ_2020__i.d.F._14.07.2023_gltg_14.07.2023_.pdf (abgerufen am 14.07.2023).

14 Siehe Modulübersicht, online unter https://www.uni-speyer.de/fileadmin/Studium/MAPA/2021-10-25_MAPA_Modulhandbuch.pdf (abgerufen am: 14.07.2023).

15 Sophia Victoria Krebs, Wolfgang Lukas, Rüdiger Nutt-Kofoth, Anja Platz-Schliebs, Karl-Heinrich Schmidt, Uwe Stadler, Eric W. Steinhauer, Editions- und Dokumentwissenschaft an der Bergischen Universität Wuppertal, in: Bibliotheksdienst 52 (2018) 12, S. 875–889. Zum Vergleich auch: Wolfgang Lukas: Was ist das Digitalisierungsinteresse der geistes- und kulturwissenschaftlichen Forschung?, in: Claudia Kauertz (Red.), Digital und analog – die beiden Archivwelten. 46. Rheinischer Archivtag 21.–22. Juni 2012 in Ratingen. Bonn 2013, S. 32–47; Sahle, Blumtritt (wie Anm. 3), S. 20–21.

16 Siehe hierzu die Informationen auf den Seiten des Studiengangs bei der Fakultät für Geistes- und Kulturwissenschaften sowie die Modulübersicht, online unter <https://www.edw.uni-wuppertal.de/de/home.html>, https://www.zpa.uni-wuppertal.de/fileadmin/zpa/Studiengaenge/Master/Master-1-Fach-Studiengaenge/Editions-_und_Dokumentwissenschaft/PO_2015_Leistungsbescheinigungen/MA-2015_EDW_-_Modul%3%BCbersicht_-_mit_Pr%3%BCfungsnummern.pdf (abgerufen am: 14.07.2023).

17 „Online-Fortbildung 4_2021: Auf dem Weg zum Digitalen Archiv – An- und Herausforderungen bei der Onlinepräsenz“ am 14.07.2023; online unter: <https://www.vda.archiv.net/online-fortbildungen/vergangene-veranstaltungen/4-2021-und-6-2021.html> (abgerufen am: 14.07.2023).

kursen für sog. „Seiteneinsteiger/innen“ an. 2019 wurden der „Praxis der Schriftgutverwaltung“ ein Grundkurs und ein Aufbaukurs „Archivische Beratung bei der Schriftgutverwaltung“, außerdem fünf Erweiterungskurse der digitalen Archivierung gewidmet.

Grundkurse ... vermitteln grundlegende Fachkenntnisse für das Archivwesen und das Records Management	Aufbaukurse ... dienen der Vertiefung von archivischen Fachkenntnissen	Erweiterungskurse ... bieten Einführungen und erweiternde Kenntnisse für neue Aufgabengebiete
GK 1-1 Einführung in das Archivwesen 13	AK 11 Überlieferungsbildung – Ziele, Methoden, Verfahren 28	EK 50-1 Grundlagen der Archivierung digitaler Unterlagen 40
GK 1-2 Einführung in das Archivwesen 14	AK 12 Nachlässe: Privates Schriftgut in Archiven 29	EK 50-2 Grundlagen der Archivierung digitaler Unterlagen 41
GK 1-3 Einführung in das Archivwesen 15	AK 14 Erschließung und Bearbeitung von Sammlungsgut 30	EK 50-3 Grundlagen der Archivierung digitaler Unterlagen 42
GK 2-1 Einführung in die Ordnung und Verzeichnung des Archivguts 16	AK 22 Schäden an Archivgut erkennen, begrenzen und behandeln (in Ludwigsburg) 31	EK 56-1 Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive 43
GK 2-2 Einführung in die Ordnung und Verzeichnung des Archivguts 17	AK 23 Bestandserhaltungs- und Notfallmanagement in kleineren und mittelgroßen Archiven 32	EK 56-2 Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive 44
GK 2-3 Einführung in die Ordnung und Verzeichnung des Archivguts 18	AK 41 Archivrecht und Rechtsfragen im Archivalltag 33	EK 61-1 Digitale Bildbearbeitung im Archiv – Anwenderschulung 45
GK 3 Aufgaben und Betrieb kleiner und mittlerer Archive 19	AK 42 Urheberrecht im Archiv 34	EK 61-2 Digitale Bildbearbeitung im Archiv – Anwenderschulung 46
GK 4 Einführung in die Paläographie – 15. – 17. Jhd. 20	AK 51 Archivische Beratung bei der Schriftgutverwaltung 35	EK 62 Photoshop im Archiv 47
GK 5-1 Einführung in die Paläographie – 18. – 20. Jhd. 21		EK 66 MidosaXML – Schulung 48
GK 5-2 Einführung in die Paläographie – 18. – 20. Jhd. 22		EK 80 Datenaufbereitung für Online-Portale am Beispiel des Archivportals-D. Einführung in XML und EAD(DDb) 49
GK 5-V Vertiefung: Paläographie des 18. – 20. Jhd. 23		EK 90 Kundenservice im Archiv 50
GK 7 Praxis der Schriftgutverwaltung 24		

Abb. 1: Archivschule, Fortbildungsprogramm 2019¹⁸

2020 wurde das Programm erweitert auf drei Kurse zu Schriftgutverwaltung und E-Akte, die neben die nunmehr sechs Kurse zur digitalen Archivierung traten.

Übergreifende Aufgaben	Bestandserhaltung
GK 1101 Einführung in das Archivwesen S. 17	AK 3101 Schäden an Archivgut erkennen, begrenzen und behandeln (Münster) S. 45
GK 1102 Einführung in das Archivwesen S. 18	AK 3111 Bestandserhaltungs- und Notfallmanagement in kleineren und mittelgroßen Archiven S. 46
GK 1103 Einführung in das Archivwesen S. 19	EK 3211 Digitale Bildbearbeitung im Archiv – Anwenderschulung S. 47
GK 1104 Einführung in das Archivwesen S. 20	EK 3212 Digitale Bildbearbeitung im Archiv – Anwenderschulung S. 48
AK 1111 Aufgaben und Betrieb kleiner und mittlerer Archive S. 21	EK 3221 Photoshop im Archiv S. 49
GK 1201 Archivrecht und Rechtsfragen im Archivalltag S. 22	
GK 1202 Archivrecht und Rechtsfragen im Archivalltag S. 23	
AK 1211 Urheberrecht im Archiv S. 24	
GK 1301 Grundlagen der Archivierung digitaler Unterlagen S. 25	Bereitstellung und Vermittlung
GK 1302 Grundlagen der Archivierung digitaler Unterlagen S. 26	GK 4201 Einführung in die Ordnung und Verzeichnung des Archivguts S. 51
GK 1303 Grundlagen der Archivierung digitaler Unterlagen S. 27	GK 4202 Einführung in die Ordnung und Verzeichnung des Archivguts S. 52
AK 1311 Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive S. 28	GK 4203 Einführung in die Ordnung und Verzeichnung des Archivguts S. 53
AK 1312 Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive S. 28	GK 4204 Einführung in die Ordnung und Verzeichnung des Archivguts S. 54
AK 1313 Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive S. 30	AK 4101 Erschließung und Bearbeitung von Sammlungsgut S. 55
GK 1411 Einführung in die Paläographie - 18. - 21. Jahrhundert S. 31	EK 4111 Midosa XML – Schulung S. 56
GK 1412 Einführung in die Paläographie - 18. - 21. Jahrhundert S. 32	EK 4121 Semantische Technologien und der Einsatz von Normdaten in der archivischen Erschließung S. 57
AK 1421 Vertiefung: Paläographie des 18. - 21. Jahrhunderts S. 33	EK 4311 Datenaufbereitung für Online-Portale am Beispiel des Archivportals-D. Einführung in XML und EAD(DDb) S. 58
GK 1432 Einführung in die Paläographie - 15. - 17. Jahrhundert S. 34	
AK 1511 Kundenservice im Archiv S. 35	
Behördenberatung und Überlieferungsbildung	Bearbeitung bestimmter Archivaliengattungen
GK 2201 Grundlagen der Schriftgutverwaltung / des Records Managements S. 37	AK 5301 Nachlässe: Privates Schriftgut in Archiven S. 61
GK 2202 Grundlagen der Schriftgutverwaltung / des Records Managements S. 38	AK 5401 Archivierung audiovisueller Medien S. 62
AK 2221 Einführung in die Elektronische Akte S. 39	
AK 2222 Einführung in die Elektronische Akte S. 40	
AK 2211 Archivische Beratung bei der Schriftgutverwaltung / dem Records Management S. 41	
AK 2101 Überlieferungsbildung – Ziele, Methoden, Verfahren S. 42	Sonstiges
	Ankündigung 25. Archivwissenschaftliches Kolloquium S. 43
	Teilnahmebedingungen und organisatorische Hinweise S. 65
	Publikationen S. 68
	Veranstaltungen im Jahresüberblick S. 72

Abb. 2: Archivschule, Fortbildungsprogramm 2020¹⁹

Seit 2021 sind die Grundlagenkurse weiter ausdifferenziert worden: Es werden nun Schriftgutverwaltung und Records Management separat behandelt, zur E-Akte wurden ebenfalls (zwei) eigene Einführungs-Kursangebote eingerichtet.

18 Das Fortbildungsprogramm für 2019 wurde nach Stand vom 10.04.2021 von der Seite der Archivschule heruntergenommen. Es liegt den Autoren noch als PDF-Download vor.

19 Das Fortbildungsprogramm für 2020 wurde nach Stand vom 10.04.2021 von der Seite der Archivschule heruntergenommen. Es liegt den Autoren noch als PDF-Download vor.

Die Kurse zur digitalen Archivierung sind neben Grundlagenvermittlung vermehrt auch auf die Vermittlung praxisnaher Inhalte ausgedehnt worden. Interesse und Bedarf sind unbestritten groß.

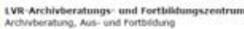
Übergeordnete Aufgaben			
GK 1101	(GK 1-1)	Einführung in das Archivwesen	S. 17
GK 1102	(GK 1-2)	Einführung in das Archivwesen	S. 18
GK 1103	(GK 1-3)	Einführung in das Archivwesen	S. 19
GK 1104	(GK 1-4)	Einführung in das Archivwesen	S. 20
GK 1201	(AK 41)	Archivrecht und Rechtsfragen im Archivaltag	S. 21
GK 1202	(AK 41)	Archivrecht und Rechtsfragen im Archivaltag	S. 22
GK 1301	(EK 50-1)	Grundlagen der Archivierung digitaler Unterlagen	S. 23
GK 1302	(EK 50-2)	Grundlagen der Archivierung digitaler Unterlagen	S. 24
GK 1303	(EK 50-3)	Grundlagen der Archivierung digitaler Unterlagen	S. 25
GK 1304	(EK 50-4)	Grundlagen der Archivierung digitaler Unterlagen	S. 26
GK 1411	(GK 5-1)	Einführung in die Paläographie - 18. - 21. Jahrhundert	S. 27
GK 1412	(GK 5-2)	Einführung in die Paläographie - 18. - 21. Jahrhundert	S. 28
AK 1311	(EK 56-1)	Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive	S. 29
AK 1312	(EK 56-2)	Archivierung digitaler Unterlagen: Pragmatische Lösungen für kleine und mittlere Archive	S. 30
AK 1211	(AK 42)	Urheberrecht im Archiv	S. 31
AK 1421	(GK 5-V)	Vertiefung: Paläographie des 18. - 21. Jahrhunderts	S. 32
AK 1111	(GK 3)	Aufgaben und Betrieb kleiner und mittlerer Archive	S. 33
Behördenberatung und Überlieferungsberatung			
GK 2201	(GK 7)	Praxis der Schriftgutverwaltung	S. 37
GK 2221		Schriftgutverwaltung und Elektronische Akte	S. 38
AK 2211	(AK 51)	Archivische Beratung bei der Schriftgutverwaltung	S. 39
AK 2101	(AK 11)	Überlieferungsberatung – Ziele, Methoden, Verfahren	S. 40
Bestandserhaltung			
AK 3101	(AK 22)	Schäden an Archivgut erkennen, begrenzen und behandeln (Ludwigsburg)	S. 43
AK 3111	(AK 23)	Bestandserhaltungs- und Notfallmanagement in kleineren und mittelgroßen Archiven	S. 44
EK 3211	(EK 61-1)	Digitale Bildbearbeitung im Archiv – Anwenderschulung	S. 45
EK 3212	(EK 61-2)	Digitale Bildbearbeitung im Archiv – Anwenderschulung	S. 46
EK 3221	(EK 62)	Photoshop im Archiv	S. 47
Bereitstellung und Vermittlung			
GK 4201	(GK 2-1)	Einführung in die Ordnung und Verzeichnung des Archivguts	S. 49
GK 4202	(GK 2-2)	Einführung in die Ordnung und Verzeichnung des Archivguts	S. 50
GK 4203	(GK 2-3)	Einführung in die Ordnung und Verzeichnung des Archivguts	S. 51
GK 4204	(GK 2-4)	Einführung in die Ordnung und Verzeichnung des Archivguts	S. 52
AK 4101	(AK 14)	Erschließung und Bearbeitung von Sammlungsgut	S. 53
EK 4111	(EK 66)	Midosa XML – Schulung	S. 54
AK 1511	(EK 90)	Kundenservice im Archiv	S. 55
EK 4311	(EK 80)	Datenaufbereitung für Online-Portale am Beispiel des Archivportals-D. Einführung in XML und EAD(DDB)	S. 56
Bearbeitung bestimmter Archivalienkategorien			
GK 5101	(GK 8)	Einführung in die Aktenkunde	S. 59
AK 5201	(AK 15)	Personenstandsunterlagen in Kommunalarchiven – Übernahme und Benutzung	S. 60
AK 5301	(AK 12)	Nachlässe: Privates Schriftgut in Archiven	S. 61
Sonstiges			
		Ankündigung 25. Archivwissenschaftliches Kolloquium	S. 34
		Teilnahmebedingungen und organisatorische Hinweise	S. 64
		Publikationen	S. 67
		Veranstaltungen im Jahresüberblick	S. 72

Abb. 3: Archivschule, Fortbildungsprogramm 2021²⁰

Das LVR-AFZ bietet bereits seit knapp zehn Jahren regelmäßig Veranstaltungen zu digitaler Archivierung und Records Management an. Ein Schwerpunkt lag und liegt dabei aus nachvollziehbaren Gründen auf den Angeboten des „Digitalen Archivs NRW“. Zuletzt wurde im November 2020 eine neu konzipierte virtuelle Fortbildungsveranstaltung zum Records Management angeboten, bei der die Zahl der Anmeldungen die zur Verfügung stehenden Plätze bei Weitem überschritten hat. Für die Fortbildungen beim LVR-AFZ werden wie bei der Archivschule Marburg häufig nicht nur Expert/innen aus dem eigenen Haus, sondern auch externe Fachleute für Vorträge, Übungen, Fragerunden eingeladen; so auch bei der Veranstaltung 2020. Diese referierten nicht nur zu den Standards der Schriftgutverwaltung oder des Records Management, wie der ISO 15489, der Erstellung eines Aktenplans, der Logik Akte-Vorgang-Dokument und dem Lebenszyklus (bzw. zum *records continuum*)²¹, sondern auch zu den Herausforderungen aus ihrem eigenen Arbeitsalltag.

20 Online unter: https://www.archivschule.de/uploads/Fortbildung/2021/Fobi-Programm_Homepage_Korr.pdf (abgerufen am: 14.07.2023).

21 Records Management und klassische Schriftgutverwaltung haben trotz vieler Schnittmengen auch einige Unterschiede aufzuweisen, die bewusst bleiben müssen. Siehe an dieser Stelle: Lemma „Lebenszyklus“ (2015), in: Archivwissenschaftliche Terminologie; Lemma „Records Continuum“ (2013), in: ebd.; Lemma „Records Management“ (2013), in: ebd.; Lemma „Schriftgutverwaltung“ (2013), in: ebd., online unter: <https://www.archivschule.de/uploads/Forschung/ArchivwissenschaftlicheTerminologie/Terminologie.html> (abgerufen am: 14.07.2023).



LVR-Archivberatungs- und Fortbildungszentrum
Archivberatung, Aus- und Fortbildung



LVR
Qualität für Menschen



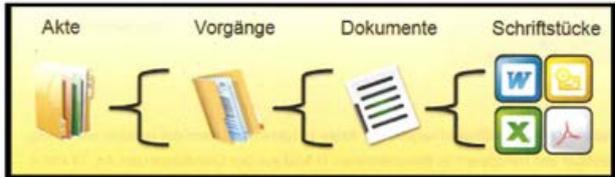
Hünxe
VERANPET

Kernfragen des Vortrags

- Was ist Schriftgutverwaltung, was ist Records Management?
- Warum ist das wichtig?
- Einige zentrale Erläuterungen
- Was hat das mit dem Archiv zu tun?
- Einige aktuelle Normen und Richtlinien
- Welchen Nutzen haben eigentlich Aktenpläne?
- Und wenn es jetzt doch alles nur noch digital gibt?

Aufbau Akte

Aufbau einer Akte in der eGoveSuite



M. Senk, 12/2020




Normative Anforderungen



3. Dezember 2020



Folie 6



LVR-Archivberatungs- und Fortbildungszentrum

Analoge und digitale Aktenführung - Records Management für Archivar*innen

Linkliste

Records Management allgemein

- Bundesentwicklungsplan (BEP) - Schriftgutverwaltung in Bundesbehörden - Einführung in die Praxis („Klemer Hoffmann“)
<https://archiv.de/Details/Hoffmann-2005-schriftgutverwaltung-in-bundesbehoe...>
- Stadtarchiv Mannheim: Akte, Vorgang und Vermerk. Ein kurzer Leitfaden zur Vorgehensführung und Schriftgutverwaltung
<https://www.mbr.ch/aktuelle/News/2018-11-14/Archiv-Akte-Vorgang-Vermerk-2018-11-14.pdf>
- Stadtarchiv Hamburg: Management analoger und digitaler Aufzeichnungen. Records Management. Ein Handbuch für die hamburgische Verwaltung
<https://www.hamburg.de/content.../management-2018-01-01.pdf>
- Arbeitsgruppe IT-gestützte Verwaltungsarbeit: Grundtatpapier „Aktuelle Relevanz von Dokumenten“
<https://www.hamburg.de/Content/Download/aktuelle-relevanz-von-dokumenten-aktuelle-relevanz-von-dokumenten.pdf>



LVR
Qualität für Menschen

Abb. 4: (v.o.n.u. und v.l.n.r.) Folien und Materialien von: Matthias Senk (LVR-AFZ), Kirsten Lehmkühl (Gema Hünxe), Hendrik Mechernich (StA Frechen) und noch einmal von Hr. Senk²²

2020 wurde außerdem die virtuelle Fortbildungsreihe „Summer Lectures“ zur digitalen Archivierung beim LVR-AFZ neu eingerichtet, die nicht minder große Resonanz fand. Sie wird 2021 fortgeführt.

Vorteil dieser Angebote und Aktivitäten ist, dass Praktiker/innen theoretische Grundlagen vermitteln sowie aus der eigenen Erfahrung berichten können – gerade auf Rückfrage –, und dass gleichzeitig eine Grundlage zur Vernetzung und zum Erfahrungsaustausch für diejenigen geboten wird, denen es noch an Praxiserfahrung fehlt.

Demgegenüber sind die Themen Records Management und digitale Archivierung in der staatlichen und kommunalen Archivausbildung ein fester, nicht wegzudenkender Bestandteil und modular mit sowohl historischen als auch aktuellen Komponenten verknüpft. Im Gegensatz zu der skizzierten Lage, kommt es hier nicht auf institutionelles oder gar individuelles Engagement an.

2.2 Die archivarische Fachausbildung

Um die hier relevanten Inhalte richtig einordnen zu können, bedarf es zunächst eines Blickes auf die nicht ganz unkomplizierte Struktur der Ausbildung von Archivar/innen in Deutschland. Lässt man die Programme der Stufe 3 des International Standard Classification of Education (ISCED) außen vor, das heißt in Deutschland insbesondere die sehr gut angenommene, für die Archive überaus wertvolle Ausbildung von Fachangestellten für Medien- und Informationsdienste, so bleibt es bei einem im europäischen Vergleich bemerkenswerten Dualismus bzw. eigentlich sogar einer Dreiteilung. Die Eigenarten und Besonderheiten dieses Modells hat Katharina Tiemann (LWL-Archivamt für Westfalen) vor

²² Wir danken den Referierenden der genannten Fortbildung an dieser Stelle herzlich für ihre Einwilligung, die Folien verwenden zu dürfen.

zwei Jahren für die Zeitschrift des Landkreistages Nordrhein-Westfalen prägnant und übersichtlich zusammengefasst.²³ Dem „normalen“ Bachelorstudium zum „Archivar/in B.A.“ stehen zwei berufsbegleitende Fernweiterbildungen sowie zwei verwaltungsinterne Ausbildungsgänge gegenüber.

Wie auch das „normale“ Bachelorstudium werden die berufsbegleitenden Studiengänge lediglich an einer deutschen Hochschule angeboten, der University of Applied Sciences in Potsdam. Der erste der beiden Fernstudienkurse richtet sich an Fach- und Verwaltungsangestellte, die eine Vorbildung der ISCED-Stufe 3 oder 4 mitbringen. Sie läuft über vier Jahre und ist streng genommen kein Studium. Es besteht aber die Möglichkeit, sich nach erfolgreichem Abschluss in einem abgetrennten Prüfungsverfahren in das „normale“ Studium einstufen zu lassen und die Prüfung zum Bachelor of Arts abzulegen.

Die zweite der beiden Fernweiterbildungen ist ein regulärer Masterstudiengang. Die Dauer beträgt sechs Semester. Als Zusatzstudium soll der Masterstudiengang sogenannten „Seiteneinsteiger/innen“ die Möglichkeit eröffnen, eine archivarisches Fachausbildung des höheren Dienstes zu erwerben. Er richtet sich daher an Absolvent/innen geisteswissenschaftlicher Fächer, die umfangreiche Berufserfahrung in Archiven mitbringen und für die Dauer des Studiums eine Beschäftigung in einem Archiv oder einer vergleichbaren Einrichtung nachweisen können.

Träger der verwaltungsinternen Ausbildungsgänge sind die staatlichen Archive des Bundes und der Länder. Nordrhein-Westfalen kommt eine Sonderstellung zu, da hier die Landschaftsverbände gemeinsam mit dem Landesarchiv ausbilden. Die Ausbildung zum gehobenen Dienst (Diplom-Archivar/in, ISCED-Level 6) läuft über drei Jahre; die Ausbildung zum/zur wissenschaftlichen Archivar/in (ISCED-Level 7) dauert zwei Jahre, wobei die sogenannten „Referendar/innen“ in aller Regel bereits eine geschichts-, rechts- oder verwaltungswissenschaftliche Vorbildung der ISCED-Stufe 8 oder zumindest des Levels 7 in Kombination mit umfangreicher Berufserfahrung vorweisen können. Beide verwaltungsinternen Ausbildungsgänge bestehen je zur Hälfte aus theoretischen Studien an der Archivschule Marburg sowie Praktika in den Ausbildungs- und anderen Archiven. In Bezug auf die hier zu analysierende Vermittlung von Kompetenzen zu elektronischer Schriftgutverwaltung und digitaler Langzeitarchivierung ist dabei darauf zu achten, dass Inhalte und Ablauf dieser Praktika von Bundesland zu Bundesland divergieren. Sehr grundsätzlich wird man sagen können, dass diejenigen, die in größeren Archivverwaltungen ausgebildet werden (z. B. Nordrhein-Westfalen, Baden-Württemberg), in der Regel bereits während der Praktika mehr Unterrichtseinheiten erhalten als die Anwärter/innen und Referendar/innen aus kleineren Archivverwaltungen. Letztere werden dafür stärker in die „alltägliche“ Arbeit im Archiv eingebunden, was umfangreiche Projekte mit Bezug zu den hier relevanten Themen beinhalten kann, aber nicht zwangsläufig muss.

Allen Studiengängen sowohl in Potsdam als auch im Rahmen der verwaltungsinternen Ausbildung ist gemein, dass die Vermittlung von Kompetenzen heute mindestens ebenso wichtig ist wie die Vermittlung von Fachwissen. Diese Ausgangsinformation ist insofern wichtig, als die Aussagen „Das habe ich nie wieder gebraucht“ oder „Die Details habe ich morgen doch sowieso wieder vergessen“ immer wieder vorgebracht werden, dem Anliegen einer kompetenzbasierten Ausbildung aber nur bedingt gerecht werden.

2.2.1 „Normales“ Bachelorstudium an der Fachhochschule Potsdam

Die Zugangsvoraussetzungen entsprechen den in Deutschland üblichen Vorgaben, das heißt Kandidat/innen müssen die Allgemeine Hochschulreife, die Fachhochschulreife oder eine als gleichwertig anerkannte berufspraktische Vorbildung mitbringen. Ablauf und Inhalte sind im Zuge einer Neuakkreditierung zuletzt zum Wintersemester 2020/21 überarbeitet und aktualisiert worden. Im Laufe des siebensemestrigen Studiums sind neun studiengangübergreifende – das heißt auch von Studierenden der Studiengänge Bibliothekswissenschaft sowie Informations- und Datenmanagement zu absolvierende – sogenannte G-Module zu besuchen:

Die ersten beiden dieser Module haben einen Workload von 120 bzw. 150 Stunden und heißen „Informationswissenschaftliche Werkstatt“ und „Einführung in die Informationsgesellschaft“. Da hier grundlegende Fach- und Sachkompetenz vermittelt wird, wird man ihnen eine große Bedeutung für eine spätere, intensivere Beschäftigung mit „elektronischen“ Fragen beimessen dürfen. Ähnliches gilt für die Module „Grundlagen Medien“ und „Information und Gesellschaft“, welche zumindest Berührungspunkte mit Grundlagen elektronischer Schriftgutverwaltung und digitaler Langzeitarchivierung aufweisen. Exemplarisch sei auf „Parameter der digitalen Reproduktion (Text, Bild), Scannen, OCR“ verwiesen. Insgesamt machen diese Schnittmengen aber nur einen Teil der insgesamt jeweils vorgesehenen 150 Stunden aus. Das Gegenteil gilt für das Modul „Webtechnologie & Informationssystem“. Die hierfür vorgesehenen 150 Stunden sind sicherlich dazu geeignet, die Basis für umfangreiche digitale Kompetenzen zu legen. Keine spezifisch digitalen Kompetenzen werden in den Modulen „Wissenschaftliches Arbeiten und statistische Methoden“ (150 Stunden), „Fachenglisch“ (120 Stunden), „Management und Recht“ (330 Stunden) und „Vertiefung Schlüsselqualifikationen“ (120 Stunden) vermittelt.

23 Katharina Tiemann, Professionalität = Fachlichkeit! Ausbildung von Fachkräften, in: EILDienst. Monatszeitschrift des Landkreistages Nordrhein-Westfalen, 4/2019, S. 254-255.

Insgesamt ist somit zu konstatieren, dass allein durch die gemeinsamen Veranstaltungen mit angehenden Bibliothekar/innen und Informationswissenschaftler/innen „digitalen“ Kompetenzen ein hoher Stellenwert eingeräumt wird. Ebenso klar ist aber, dass diese Kompetenzen aufgrund der breit angelegten Zielgruppe noch nicht archivspezifisch im eigentlichen Sinne des Wortes sind.

Von den neun Pflichtmodulen für Archivar/innen, den sogenannten A-Modulen, sind „Erschließung von Archivgut“ (240 Stunden), „Handschriften des 19./20. Jahrhunderts“ (150 Stunden), „Geschichtswissenschaftliche Grundlagen“ (330 Stunden), „Verwaltungsgeschichte und Archivrecht“ (240 Stunden), „Marketingstrategien & Öffentlichkeitsarbeit in Archiven“ (210 Stunden), „Archivwissenschaft II“ (210 Stunden) und „Archivbau und Benutzung“ (210 Stunden) bestenfalls am Rande für das hier in Rede stehende Thema von Interesse. „Digitale“ Kompetenzen werden gebündelt in den Modulen „Archivwissenschaft I“ (240 Stunden) und „Digitale Präsentation von Archivalien“ (210 Stunden). Der Schwerpunkt von „Archivwissenschaft I“ liegt dabei u. a. auf elektronischem Records Management und eGovernment sowie Fragen der digitalen Archivierung. Bei der „Digitalen Präsentation von Archivalien“ steht neben der Digitalisierung von papiergebundenen Vorlagen, also der Digitalisierung im engeren Sinne, mit digitalen Editionen ein Thema im Mittelpunkt, das zumindest im kommunalarchivarischen Arbeitsalltag bislang nur eine nachgeordnete Rolle spielt.

Alles in allem wird den hier zu analysierenden Fragen somit eine hohe Bedeutung beigemessen. Es ist sowohl semantisch als auch programmatisch ein Signal, wenn das Modul „Archivwissenschaft I“ einen deutlich erkennbaren digitalen Schwerpunkt aufweist, es zugleich aber keine Module „Hilfswissenschaften“ oder „Überlieferungsbildung“ gibt.

Im dritten und vierten Semester haben die Studierenden die Möglichkeit, zwei von vier Wahlpflichtmodulen zu absolvieren, die jeweils mit 270 Stunden hinterlegt sind. Modul 4 „Digitale Unterlagen“ darf in diesem Zusammenhang mit Fug und Recht als „Spezialisierung“ auf den digitalen Bereich gelten. Diese Spezialisierung kann jede/r Studierende im sechsten Semester weiter ausbauen: zum einen durch die Wahl eines entsprechenden Themas für die Abschlussarbeit, zum anderen im Zuge der Wahl zwischen dem integrativen – das heißt auch hier wieder studiengangübergreifenden – Modul „Information Visualization“ und „Vertiefung Hilfswissenschaften“.

Alles in allem weist das „normale“ Bachelorstudium somit einen deutlich erkennbaren Schwerpunkt in „modernen“, informationswissenschaftlichen Feldern auf. Fraglich erscheint in Bezug auf die Erfahrung bei Vorstellungsgesprächen u. ä. aber, inwiefern es den Studierenden möglich ist, die vermittelten theoretischen Kenntnisse durch eigene praktische Erfahrungen zu vertiefen, bzw. besser: abzusichern. Im Praxissemester steht – aus sehr gut nachvollziehbaren Gründen – eine klassische Verzeichnungsübung im Zentrum des Interesses. Das sechswöchige Praktikum im zweiten Semester und das Projekt im sechsten Semester reichen in der Regel wohl kaum aus, um umfangreiche Einblicke in die Berufspraxis zu erhalten.

2.2.2 Berufsbegleitendes Masterstudium an der Fachhochschule Potsdam

Mangelnde Berührungspunkte mit dem Arbeitsalltag sind für diejenigen, die mit mindestens einjähriger Berufserfahrung vor Studienbeginn und mit einer mindestens für die Dauer des Studiums befristeten Beschäftigung in einem Archiv in die Ausbildung starten, kein Problem. Schwierigkeiten bereitet den Betroffenen gelegentlich aber durchaus der Zuschnitt der Stelle. Wer auf einer nicht dem höheren oder gehobenen Dienst zugeordneten Stelle „sitzt“, dem fehlt es erfahrungsgemäß häufig an Möglichkeiten, die von der Fachhochschule intendierte Verbindung von Studieninhalten und Arbeitsalltag zu bewerkstelligen.²⁴

Blickt man auf die Inhalte, so besteht das berufsbegleitende Masterstudium aus 16 Pflichtmodulen à 150 Stunden Workload. Zwei von diesen Modulen sind an dieser Stelle von höchstem Interesse: Sie heißen nicht umsonst „Digitale Archive“ und „Digitale Archivierung“. Das Modul „Information und Gesellschaft“ und das Modul „Records Management“ weisen hingegen – zumindest, wenn man sich an den Modulhandbüchern orientiert – weniger Schnittmengen mit den hier zu analysierenden Fragestellungen auf als die entsprechenden Module des „normalen“ Bachelorstudiengangs. Ein Gespräch mit Absolvent/innen beider Kurse führt zwar zu dem Ergebnis, dass diese Diskrepanz zwischen beiden Studiengängen in der Realität deutlich schwächer ausgeprägt ist. Insgesamt bleibt es aber bei dem Eindruck, dass den klassischen analogen Fachaufgaben im berufsbegleitenden Masterstudiengang größere Bedeutung beigemessen wird als im Bachelorstudiengang. Zumindest in Bezug auf die nichtstaatlichen Archive im Rheinland entspricht diese Verteilung auch nach wie vor eher Anforderungen des Arbeitsalltags in kleinen und mittelgroßen Einrichtungen. Dennoch ist der Befund etwas überraschend, da insbesondere im Bereich der Hilfswissenschaften und der analogen Archivarbeit bei Kandidat/innen mit abgeschlossenem (Geschichts-)Studium und mindestens einjähriger Berufserfahrung deutlich bessere Vorkenntnisse zu vermuten sind als bei Studienanfänger/innen im eigentlichen Sinne des Wortes.

Als sehr gelungen darf die Gestaltung des Wahlpflichtbereiches im vierten Semester bezeichnet werden. Bei der Auswahl von zwei aus sechs Wahlpflichtmodulen kann jede/r Studierende zwischen einem Schwerpunkt im Bereich der elektronischen Archivierung bzw. ihren Nachbardisziplinen (IuK-Technologien, Datenbankarchivierung, Metabeschreibungssprachen)

24 Als geradezu paradisiisch und zur Nachahmung empfohlen – wenn das Eigenlob an dieser Stelle erlaubt ist – darf hingegen die Verbindung von Volontariat und Fernstudium gelten, wie sie im LVR-AFZ seit einigen Jahren mit sehr großem Erfolg praktiziert wird.

oder aber im Bereich der klassischen historisch-organisatorischen Archivarbeit (Managementtechniken, wissenschaftliche Editionen, Geschichtsprjekt) wählen. Auch „Mischformen“ sind selbstverständlich möglich. Abschließend bietet dann auch die Wahl des Themas für die Masterarbeit eine Möglichkeit zur Spezialisierung. Zumindest für das Rheinland ist diesbezüglich – wenn auch bei Weitem nicht repräsentativ – zu konstatieren, dass Fragen der elektronischen Archivierung bzw. des elektronischen Records Management zunehmend an Bedeutung gewinnen. Wo es gelingt, das Ansinnen des berufsbegleitenden Studiums in die Arbeitswirklichkeit zu transponieren, bieten sowohl Abschluss- als auch Hausarbeiten mannigfaltige Möglichkeiten, „digitale“ Projekte voranzutreiben oder anzustoßen. Exemplarisch sei auf die Arbeitsergebnisse unserer beiden Volontär/innen – Dr. Henrike Bolte (heute: StA Dortmund) und Hendrik Mechernich M.A. (heute: StA Frechen) –, aber auch die 2021 abgeschlossene Studie einer Kollegin aus dem Kreis Euskirchen verwiesen.

2.2.3 Fernweiterbildung Archiv an der Fachhochschule Potsdam

Insgesamt werden in vier Jahren 27 Pflicht- und zwei (von vier zur Auswahl stehenden) Wahlpflichtmodule belegt, zu denen es jeweils einen Einstieg und eine Zusammenführung in Präsenz gibt. Sechs der 27 Pflichtmodule („Kulturelles Erbe im digitalen Zeitalter“, „Webtechnologie und Informationssysteme“, „Digitale Aspekte der Archivwissenschaft“, „Digitale Archivierung“, „Digitale Präsentation von Archivalien“, „Records Management/eGovernment“) und eines der vier Wahlpflichtmodule („Datenbankarchivierung“) sind dem „digitalen“ Arbeitsfeld zuzuordnen. Ihr Anteil liegt somit zwischen 20 und 25 Prozent. Da davon auszugehen ist, dass beim Aufstieg von ISCED-Stufe 4 zu Stufe 5 der Vermittlung von grundlegenden archiv-, geschichts- und hilfswissenschaftlichen Kenntnissen große Bedeutung zukommt, ist dies ein ausgesprochen hoher Wert. Besonders aussagekräftig ist die Prozentangabe aber dennoch nicht. Der Umfang der zu erbringenden Prüfungsleistungen und somit der von den Teilnehmenden zu erbringende Aufwand können laut mündlicher Auskunft stark divergieren: Sie reichen von längeren Hausarbeiten bis zu kleineren Hausaufgaben. Ein großer Vorteil ist vergleichbar zum Masterstudiengang die Möglichkeit zur Verbindung zwischen Arbeitsalltag und Prüfungsleistungen; die Schwierigkeit, die eigene Stellenbeschreibung mit den Inhalten des Studiums in Einklang zu bringen, ist bei Beschäftigten des mittleren Dienstes in der Regel jedoch noch größer als bei Absolvent/innen des Masterstudiengangs. Erfahrungsgemäß bedarf es verbindlicher Absprachen zwischen Vorgesetzten und Fachangestellten, um eine sinnvolle Verknüpfung zu gestalten. Wo eine solche Verknüpfung gelingt, ist sie in der Regel aber nicht nur für die Teilnehmenden von Vorteil: Wenn sie gemeinsam diskutiert und sinnvoll angelegt sind, sind Prüfungsleistungen nachnutzbar und es entsteht eine klassische „win-win-Situation“. Um es ganz plastisch auszudrücken: Eine Teilnehmende, die von ihrem Arbeitgeber Freiräume bekommt, bei der Themenwahl unterstützt und begleitet wird, arbeitet ohne Bedenken in ihrer Freizeit an einem dienstlich verwertbaren Thema. Eine im öffentlichen Dienst aus gutem Grund eigentlich nicht vorgesehene Situation wird ausnahmsweise und plötzlich zu einem selbstverständlichen Nutzen.

2.2.4 Gehobener Dienst, verwaltungsintern

Der große Vorteil der verwaltungsinternen Ausbildung ist sicherlich der hohe Anteil von Praktika bzw. die damit einhergehende Möglichkeit, Auszubildende so praxisnah wie möglich zu beschäftigen. In Nordrhein-Westfalen absolvieren die Archivinspektoranwärter/innen abgesehen von zwei mehrmonatigen „Stages“ in ihren Ausbildungsarchiven (LAV NRW, LWL-Archivamt, LVR-AFZ) zwei weitere Praktika in Archiven anderer Sparten. Während all dieser Stationen sollen sie idealerweise auch mit Fragen der digitalen Langzeitarchivierung und der (elektronischen) Schriftgutverwaltung in Berührung kommen und die Gelegenheit erhalten, Diskrepanzen, aber auch Übereinstimmungen zwischen Theorie und Praxis „am eigenen Leib“ zu erfahren.

Blickt man auf das Curriculum der Archivschule Marburg, so scheinen die hier relevanten Kompetenzen weniger konzentriert auf als bei Durchsicht der Modulhandbücher der Fachhochschule Potsdam. Die Regel sind vielmehr Module, welche auch Berührungspunkte mit dem Feld der digitalen Archivierung bzw. der elektronischen Schriftgutverwaltung erkennen lassen, ohne sich ausschließlich „digitalen“ Themen zu widmen. Von den 116 für das Modul „Archivische Bearbeitung besonderer Überlieferungsformen“ vorgesehenen Stunden entfällt etwas mehr als die Hälfte auf die Fächer „Digitale Langzeitarchivierung“ (35 Stunden) und „Archivierung audiovisueller Medien“ (28 Stunden). Das Modul „Archivfachliche Querschnittsaufgaben“ nimmt insgesamt 155 Stunden in Anspruch. Hiervon entfallen etwas mehr als ein Drittel auf „IT-Fächer“: 50 Stunden sind für „Einführung in die Fach-IT“, sieben für „Reprographie und Konversionsformen“ vorgesehen. Im Modul „Einführung in archivverwandte Arbeitsbereiche“ sind die Hälfte der insgesamt 54 Stunden für eine „Einführung in Informations- und Datenmanagement“ vorgesehen. Das 95-stündige Modul „Vorgangsbearbeitung in der Verwaltung“ beinhaltet 28 Stunden für das Thema „Elektronische Verfahren und Vorgangsbearbeitung in der Verwaltung“. Darüber hinaus werden elektronische Themen implizit oder explizit im Rahmen des Moduls „Grundlagen der Archivwissenschaft“ (100 Stunden) sowie den Exkursionen (60 Stunden) thematisiert. Auch wenn man in Rechnung stellt, dass Workload der lege artis modularisierten Studiengänge und die hier angegebene Zahl von Unterrichtsstunden nur bedingt miteinander vergleichbar sind, fällt doch auf, dass in Marburg ein Schwerpunkt der

Ausbildung auf geschichts-, hilfs-, und archivwissenschaftlichen, aber auch auf philologischen Grundlagenkenntnissen und Kernkompetenzen liegt. Diese Entwicklung kann man durchaus kritisieren. Wenn im Modul „Vorgangsbearbeitung in der Verwaltung“ mehr als doppelt so viel Zeit auf die „Analyse von Akten und Geschäftsgängen“ als für den Umgang mit Dokumentenmanagementsystemen veranschlagt wird, so ist dies zumindest diskussionswürdig.

Bei aller Kritik sollte und darf die eigentliche Zielgruppe des Studiums in Marburg aber nicht aus dem Auge verloren werden. Auch wenn sich unter den Inspektoranwärter/innen seit Jahren immer wieder zahlreiche Personen mit Bachelor- und Mastertitel finden, geht es grundsätzlich darum, Studienanfänger/innen mit Abitur oder Fachhochschulreife für den Archivdienst zu befähigen. Insbesondere mit Blick auf die Herausforderungen, die Absolvent/innen in nichtstaatlichen Archiven erwarten, schließt dies in jedem Fall vertiefte geschichts-, grund- und archivwissenschaftliche Kenntnisse ein. Man mag über den Anteil von Aktenkunde, Latein, Französisch und „kleinen“ Hilfswissenschaften streiten und diesen perspektivisch reduzieren: Der nach wie vor vorhandene geschichts- und hilfswissenschaftliche Schwerpunkt scheint im Vergleich zu den recht unspezifischen G-Modulen des „normalen“ Bachelorstudiengangs in Potsdam aber eher ein Vor- als ein Nachteil zu sein. Eine Reduzierung wäre nur dann sinnvoll, wenn in den Kursen tatsächlich nur noch Absolvent/innen säßen, die mindestens einen Bachelor in Geschichtswissenschaft mitbrächten. Ob dies wirklich wünschenswert, geschweige denn rechtlich möglich wäre, erscheint mehr als ungewiss. Abgesehen von punktuellen Verschiebungen sind daher eher die 18-monatigen Praktika der Platz, um entsprechende Inhalte zu vermitteln. Es wäre sinnvoll und hilfreich, wenn Kenntnisse zu Dokumentenmanagementsystemen, DA NRW, DIMAG, SORI, TheLMA usw. nicht im Unterricht, sondern „on the job“ vermittelt würden. Hierzu bedürfte es aber sicherlich einer nochmals verstärkten Abstimmung zwischen den einzelnen Archivverwaltungen und vielleicht auch des Mutes, eingefahrene Abläufe infrage zu stellen, die manchmal vielleicht eher dazu dienen, die Arbeitsbelastung, die Ausbildung im laufenden Dienstbetrieb zwangsläufig mit sich bringt, zu reduzieren.

2.2.5 Höherer Dienst, verwaltungsintern

Die für den gehobenen Dienst noch anstehende Anpassung an den „Bologna-Prozess“ ist für den höheren Dienst bereits vollzogen worden. Dennoch ergibt sich ein ähnlicher Befund, was daraus schließen lässt, dass die skizzierte Verteilung zwischen „digitalen“ und „analogen“ Kompetenzen und Kenntnissen zu Beginn der 2010er-Jahre noch als sinnvoll erachtet wurde. Im Modul „Archivwissenschaft und Fachinformatik“ nehmen die Fächer „Einführung in die fachbezogene Informatik – IT Technologie“, „Einführung in die fachbezogene Informatik – XML“ und „Begutachtung von IT-Anwendungen für Archive“ knapp die Hälfte der insgesamt 105 Stunden in Anspruch. Von den 175 für „Archivarische Fachaufgaben“ vorgesehenen Stunden entfallen 42 auf „Digitale Langzeitarchivierung“ und „Reprographie und Konversionsformen“; die Fächer „Archivische Überlieferungsbildung bei Sammlungsgut und AV-Medien“ respektive „Erschließung von Sammlungsgut und AV-Medien“ lassen ebenso wie die Exkursionen (90 Stunden) zumindest Überschneidungen erkennen. Für das Modul „Records Management“ sind 84 Stunden vorgesehen, die zum größten Teil auch Bezüge zum elektronischen Bereich erkennen lassen.

Selbst wenn man andere „nicht digitale“ archivwissenschaftliche Fächer und Module ausklammert, so überrascht es doch, dass dem nicht weniger als 420 Stunden für die Module „Historische Hilfswissenschaften“, „Mittelalterliche und frühneuzeitliche Schriftquellen“ sowie „Schriftgut des 19. bis 21. Jahrhunderts“ gegenüberstehen. Es ist sehr gut nachvollziehbar – der berufsbegleitende Masterstudiengang in Potsdam sieht mindestens 450 Stunden für Hilfswissenschaften vor –, dass, angesichts des seit Jahrzehnten sinkenden Stellenwerts hilfswissenschaftlicher Kenntnisse an Universitäten, Raum und Zeit vorhanden sein muss, um Archivar/innen zu befähigen, „ihre“ Quellen überhaupt noch lesen zu können. Dennoch muss die Frage erlaubt sein, ob Referendar/innen, die in der Regel eine Promotion in (Rechts-) Geschichte mitbringen, umfangreicher Einführungen in Verfassungs-, Verwaltungs- und Rechtsgeschichte bedürfen. Es erscheint zumindest möglich, dass diese Stunden besser und ertragreicher in praktische Übungen zum Umgang mit Ingesttools, Lösungen zur elektronischen Langzeitarchivierung usw. eingesetzt würden. Letzteres bedürfte aber natürlich der Bereitschaft von Entwickler/innen und Archivverwaltungen, „ihren“ Referendar/innen in Marburg, aber eben auch deren Kolleg/innen aus anderen Bundesländern sowie Studierenden in Potsdam, Testsysteme zur Verfügung zu stellen. Davon abgesehen gilt für die Referendar/innen das Gleiche wie für die Anwärter/innen: Die praktische Ausbildung sollte und muss der Ort sein, um möglichst viele elektronische Hilfsmittel und Lösungen nicht nur gezeigt zu bekommen, sondern auch damit zu arbeiten. Die Vielzahl von Abschlussarbeiten, die in den letzten Jahren zu Themen der digitalen Langzeitarchivierung und des elektronischen Records Management entstanden sind, spricht dafür, dass es den Referendar/innen weder an Interesse noch an Kenntnissen mangelt. Zu wünschen wäre allerdings, dass mehr Archivverwaltungen dem Vorbild des Landesarchivs Baden-Württemberg und der Fachhochschule Potsdam folgen. Wenn es die Qualität zulässt, sollten die Arbeiten der Öffentlichkeit zugänglich gemacht werden. Transparenz, Vernetzung und Teilhabe sind nicht nur für das Gelingen von Ausbildung, sondern auch von elektronischer Archivierung von zentraler Bedeutung. So wie unsere Volontär/innen nach Möglichkeit voll in die alltägliche Arbeit integriert werden, sollte dies auch bei Referendar/innen, Anwärter/innen und Absolvent/innen der Fernweiterbildung angestrebt werden.

3 Fazit

Ungeachtet aller Kritik im Detail, sind die beiden Ausbildungsstätten in Marburg und Potsdam von zentraler Bedeutung, um Wissen zu Records Management und digitaler Archivierung zu vermitteln. Sie sind nicht als konkurrierende, sondern vielmehr als sich essenziell ergänzende Angebote zu verstehen, ohne die Beruf und Fach nicht bestehen könnten.

Blickt man auf die Inhalte, so wäre es erstrebenswert, Referendar/innen, Anwärter/innen und künftige Absolvent/innen der Fernstudiengänge (noch) stärker in den Arbeitsalltag der Ausbildungsarchive zu integrieren. Weniger wichtig als Verschiebungen im Curriculum ist es, Tätigkeiten in der und durch eigene Praxis kennenzulernen.

Gleichzeitig sollten neben dem Mut zur Fort- und Weiterbildung auch der noch umfassendere Austausch und die stärkere Vernetzung vorangetrieben werden: Mit allen Vor- und Nachteilen, die aufgrund unterschiedlicher Kenntnis- und Kompetenzstände hervorgerufen werden.

Nicht zuletzt muss das hohe Niveau von Fort- und Weiterbildungsangeboten beibehalten und stetig weiterentwickelt werden. Gerade weil klassische Universitäten und Verwaltungsfachhochschulen nur wenige entsprechende Angebote machen, ist das Konzept des lebenslangen Lernens von zentraler Bedeutung. Was in der Fachausbildung der Archivar/innen angelegt ist, muss im Rahmen von Fort- und Weiterbildung stetig ausgebaut und vertieft werden. „Seiteneinsteiger/innen“ muss die Möglichkeit zur Weiterqualifizierung in noch stärkerem Maße als bisher ermöglicht werden.

Es braucht mehr Mut – zur Schaffung und Nutzung der Ausbildungsbildungsmöglichkeiten, zu deren Verbesserung; zur Schaffung neuer, Erweiterung bestehender und zunehmenden Nutzung von Fortbildungsmöglichkeiten; und nicht zuletzt zur stärkeren Vernetzung, zur interdisziplinären Kooperation, und natürlich zu mehr archivfachlichem Austausch auch und gerade abseits perfekter Lösungen aus Theorie und Praxis!

Die Entwicklung einer Datenbankstruktur für die Erschließung digitaler Archivalien im Stadtarchiv Nürnberg

Annette Birkenholz, Stadtarchiv Nürnberg

Mit der bevorstehenden Einführung der Langzeitarchivierungslösung DiPS.kommunal im Jahr 2019 stellte sich für das Stadtarchiv Nürnberg die Frage, auf welche Weise das digitale Archivgut künftig erschlossen werden soll. Zu diesem Zeitpunkt nutzte das Stadtarchiv Nürnberg bereits seit 25 Jahren die Software FAUST für die Erschließung des Archivguts, aber auch für die Erfassung weiterer Informationen. Sowohl die Gesamtstruktur als auch der Aufbau der einzelnen Datenbanken, insbesondere derjenigen, die der Erschließung des Archivguts dienen, waren in dieser Zeit stetig gewachsen. So verwundert es nicht, dass die langjährigen Erfahrungen und die etablierten Strukturen den Rahmen für die Überlegungen zur Erschließung des digitalen Archivguts vorgeben. Zugleich setzten die vorhandenen Strukturen Grenzen für die Integration der Verzeichnung digitaler Archivalien in die bestehende Verzeichnungsdatenbank. Eine komplette Neukonzeption der Erschließung, ein Neuaufbau der Verzeichnungsdatenbank und eine Migration der vorhandenen Daten samt ihren Verknüpfungen hätten einen immens hohen Aufwand erfordert. Daher galt es, die Anforderungen an die Erschließung des digitalen Archivguts so gut wie möglich mit den bestehenden Strukturen zu verbinden. Die daraus hervorgegangene Lösungsidee soll in diesem Tagungsbeitrag vorgestellt werden.¹

Da FAUST eine recht individuelle Nutzung zulässt, wird zunächst die konzeptionelle Entwicklung der Erschließung mit FAUST im Stadtarchiv Nürnberg zusammengefasst, die sich auch im Lösungsansatz für die Erschließung digitaler Archivalien widerspiegelt. Zudem wird auf die Ausgangslage eingegangen, aufgrund der die Verzeichnung des digitalen Archivguts in der vorhandenen Verzeichnungsdatenbank ungünstig erschien. Im zweiten Teil wird erläutert, welche Anforderungen bezüglich der Erschließung digitaler Archivalien erfüllt werden sollten und welche Fragen hinsichtlich der Umsetzungsmöglichkeiten auftraten. Im dritten Teil wird der Lösungsvorschlag beschrieben und es wird dargestellt, inwieweit es gelungen ist, die Anforderungen umzusetzen. Anzumerken ist, dass das Augenmerk dieses Beitrags auf der allgemeinen Konzeption der FAUST-Erschließungsdatenbanken liegt. Die Konzeption der für die verschiedenen Archivaliengattungen benötigten Erfassungsmasken wird nicht thematisiert. Außerdem wurden zahlreiche grundsätzliche Überlegungen angestellt, beispielsweise zur Bestands- und Signaturbildung sowie zum Umgang mit hybridem Archivgut und Repräsentationen, auf die hier nicht detailliert eingegangen werden kann.

1 Die Entwicklung der Erschließung mit FAUST im Stadtarchiv Nürnberg

Mitte der 1990er-Jahre wurde im Stadtarchiv Nürnberg begonnen, das Archivgut mittels der Erschließungssoftware FAUST zu erfassen.² Anfangs diente FAUST lediglich der Erstellung von Findbüchern. Datenbankrecherchen waren aufgrund der zu dieser Zeit noch beschränkten technischen Ausstattung zunächst nicht im Blick. Vor diesem Hintergrund fehlte ein Konzept für die Datenbanknutzung, was sich darin zeigte, dass zahlreiche Datenbanken angelegt wurden, die sich an den Archivbeständen orientierten.³ Die entstehende Unübersichtlichkeit sowie Nachteile für die Datenbankadministration erforderten die Überarbeitung und Zusammenführung von Datenbanken. Dies mündete Ende des Jahres 2006 in ein Gesamtkonzept für den Einsatz von FAUST.⁴ Von besonderer Bedeutung sind dafür die Nutzung von Indexfeldern und hierarchischen Thesauri sowie Verknüpfungen des Datenbanksystems. Seitdem werden für die Erschließung des Archivguts im Wesentlichen zwei Datenbanken verwendet, die als Beständeübersicht und Beständedatenbank bezeichnet werden.⁵ Die Beständeübersicht enthält die Gliederung und Beschreibung der Bestände,

1 Bei diesem Beitrag handelt es sich um eine überarbeitete Fassung des am 22.03.2021 unter dem Titel „Der Weg des Stadtarchivs Nürnberg bei der Erschließung digitaler Archivalien in FAUST“ gehaltenen Impulsvortrags. Für die schriftliche Fassung wurden neue Screenshots angefertigt, die den Entwicklungsstand der Objektarten (Erfassungsmasken) in FAUST im April 2021 wiedergeben. Außerdem sind Ergebnisse der Diskussion des Lösungsvorschlags während der bzw. im Anschluss an die Tagung eingeflossen, ebenso die Präsentation der Lösungsidee beim DiPS.kommunal-Nutzergruppentreffen im Juni 2021. Ein Dank für den kollegialen Austausch gilt insbesondere Frau Hannah Ruff (LWL-Archivamt für Westfalen), Herrn Dr. Peter Worm (Stadtarchiv Münster) und Frau Annett Schreiber (Stadtarchiv Bochum) sowie Frau Nicola Bruns (LWL-Archivamt für Westfalen) für ihre prägnanten Formulierungen. Für weitere frühe Diskussionen sei ebenfalls Frau Schreiber sowie Frau Dr. Katharina Ernst (Stadtarchiv Stuttgart) und Herrn Prof. Dr. Christian Keitel (Landesarchiv Baden-Württemberg) gedankt.

2 Walter Bauernfeind, Erschließung des Archivguts und Datenbankrecherche, in: Michael Diefenbacher, Horst-Dieter Beyerstedt, Ulrike Swoboda, Steven M. Zahlaus (Hrsg.), 150 Jahre Stadtarchiv Nürnberg 1865–2015. Begleitband zur Ausstellung des Stadtarchivs Nürnberg, Neustadt/Aisch 2015, S. 113–122, hier S. 116f.

3 Ebd., S. 118f.

4 Ebd., S. 120.

5 Zu weiteren Datenbanken etwa für die Tiefenerschließung von Gerichtsbüchern und Personendaten, für die Bibliothek, für Spezialinventare und die Stadtchronik / Zeitgeschichtliche Sammlung vgl. ebd., S. 120–122 sowie die Abbildung zur Datenbankstruktur auf S. 119. Die Beschreibung der Struktur aus dem Jahr 2015 ist im Wesentlichen weiterhin gültig.

aber auch Module zu Akzessionen, Recherchestrategien, Bearbeitern u. a. m. Die Beständedatenbank dient dagegen zur Verzeichnung der einzelnen Archivalieneinheiten. Die Verbindung zwischen Archivalien und Tektonik ist durch ein Referenzfeld in der Beständedatenbank gegeben, welches die Bestandssignatur enthält. So kann von jeder Verzeichnungseinheit die dazugehörige Bestandsbeschreibung aufgerufen werden. In beiden Datenbanken sind für die jeweilige Ebene (Bestand oder Archivale) relevante Indices und Thesauri hinterlegt.

Die konzeptuellen Überlegungen beim Aufbau und der Ausgestaltung dieser beiden Datenbanken, aber auch die im Laufe der Zeit sehr komplex gewordenen Strukturen, stellten den Ausgangspunkt für die Überlegungen zur Erschließung des digitalen Archivguts dar. Im Jahr der Einführung von DiPS.kommunal waren in der Beständedatenbank bereits über 1,2 Mio. Objekte in über 100 Objektarten (Erschließungsmasken) erfasst. Bei zahlreichen Objekten sind aufgrund der fortschreitenden Digitalisierung des Archivguts zudem eine oder mehrere Dateien angehängt. Auch bei den Zugriffsberechtigungen bzw. der Zuteilung von Berechtigungen für die Durchführung bestimmter Aktionen mit FAUST waren Nachteile und Einschränkungen aufgrund der im Laufe der Zeit gewachsenen Strukturen zu erwarten. Trotz laufend vorgenommener Nachbesserungen am Aufbau der Datenbanken, wie etwa der Zusammenfassung von Objektarten, war offensichtlich, dass eine komplette Überarbeitung der bestehenden Datenbank notwendig wäre, um digitales Archivgut ohne Beeinträchtigungen durch überkommene Strukturen erschließen zu können. Mindestens genauso wichtig waren Überlegungen zur Performance des Systems, etwa, wenn zukünftig vermehrt automatisierte Importe und Gruppenkorrekturen zur Verzeichnung digitaler Archivalien vorgenommen werden. Beide Aspekte ließen es ratsam erscheinen, sich für den Aufbau einer zweiten Verzeichnungsdatenbank für digitales Archivgut zu entscheiden. Allerdings musste gewährleistet werden, dass diese Lösung den grundsätzlichen Anforderungen an die Erschließung von Archivgut in seinen unterschiedlichen Erscheinungsformen gerecht wird. Diese Anforderungen und die damit verbundenen Fragestellungen werden im Folgenden beschrieben.

2 Anforderungen und Fragestellungen

Aufgrund der eingangs geschilderten Erfahrungen des Stadtarchivs Nürnberg aus der Anfangszeit der datenbankgestützten Erschließung wie auch durch Fachliteratur war bewusst, dass ein einheitliches Nachweissystem für alle Archivalien und deren unterschiedliche Erscheinungsformen erforderlich ist.⁶ Die inhaltliche Recherche sowie die Verwaltung von Zugriffsrechten und Schutzfristen sollten ausschließlich über FAUST erfolgen. Das Ziel bestand im eindeutigen Nachweis aller Archivalien und ihrer Erscheinungsformen in FAUST für die Recherche, Bestellung, Bereitstellung und die Bestandserhaltung, soweit die letztgenannte Funktion nicht von DiPS.kommunal erfüllt wird. Recherchen müssen nach inhaltlichen Gesichtspunkten erfolgen können, unabhängig davon, ob es sich um analoges, digitales oder hybrides Archivgut handelt. Die Zusammenhänge von hybridem Archivgut müssen erkennbar sein und die Existenz unterschiedlicher Repräsentationen muss ersichtlich sein. Ebenso ist es notwendig, gezielt nach bestimmten Arten von Archivgut recherchieren zu können, zum Beispiel auch für statistische Zwecke. Außerdem muss bei der Erschließung eindeutig sein, wie Archivgut zu erfassen ist und welche Metadaten an welcher Stelle einzutragen sind.

Von Anfang an stand fest, dass eine einheitliche Tektonik erforderlich ist. Auch für die inhaltliche Klassifikation und die Verschlagwortung des Archivguts durfte es keinen Unterschied machen, ob Archivgut analog oder digital vorliegt. Größeren Raum nahmen auch die Überlegungen zur Signaturbildung und der fortlaufenden Zählung der Archivalien ein. Wenn analoge und digitale Archivalien eines Bestands in zwei verschiedenen Datenbanken erfasst werden, darf dies nicht zu Beeinträchtigungen bei der fortlaufenden Nummerierung führen. Es sollte kein zusätzlicher Arbeitsaufwand für das Abgleichen von Nummern entstehen und Fehler wie doppelte Signaturen waren unbedingt zu vermeiden. Außerdem war zu überlegen, auf welche Weise die Signatur für digitale Archivalien eingetragen werden soll. Die Signatur ist traditionell ein Pflichtfeld. Da nun aber die Verzeichnungsdatensätze durch den Import von Metadateien aus DiPS.kommunal angelegt werden, musste das Zusammenspiel von Feldeigenschaften, Importformaten und Gruppenkorrekturen in FAUST durchdacht sowie Abläufe festgelegt werden.

Weitere Betrachtungen galten dem Umgang mit dem bereits in FAUST verzeichneten, aber noch nicht in DiPS.kommunal übernommenen digitalen Archivgut. Insbesondere digitale Fotobestände waren schon in größerem Umfang in FAUST erfasst, aber auf Laufwerken gespeichert und als sogenannte „Digitale Dokumente“ mit den Datensätzen in FAUST verlinkt. Disketten, CD-ROMs oder USB-Sticks können als eigenständiges Archivale verzeichnet sein. Zumeist sind sie aber als Bestandteil eines Aktenordners ins Archiv gelangt und wurden auch in diesem Zusammenhang belassen. Bei der Verzeichnung wurden die Datenträger im Enthält- und Darin-Vermerk sowie in der Umfangsangabe oder im Bemerkungsfeld erfasst. Für die bereits detailliert erschlossenen und mit Signaturen versehenen Fotobestände ist es unerlässlich, dass die vorhandenen Erschließungsinformationen erhalten bleiben und auch keine Schwierigkeiten hinsichtlich

6 Christian Keitel, Das Repräsentationsmodell des Landesarchivs Baden-Württemberg, in: Susanne Wolf (Hrsg.), Neue Entwicklungen und Erfahrungen im Bereich der digitalen Archivierung. Von der Behördenberatung zum Digitalen Archiv. 14. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ vom 1. und 2. März 2010 in München, München 2010, S. 69–82, hier S. 69f.

der bereits etablierten Signaturen entstehen. Mit der Übernahme des Inhalts der Datenträger in DiPS.kommunal und der damit verbundenen Bildung von inhaltlichen Einheiten ist eine Veränderung gegenüber der bisherigen Verzeichnung unausweichlich. Jedoch muss der Zusammenhang des hybriden Archivguts nachvollziehbar bleiben.

Eine gesondert zu betrachtende Situation ergab sich zudem im Hinblick auf Sicherungsdigitalisate und die gegebenenfalls dazu vorhandenen Nutzungsrepräsentationen, die mit DiPS.kommunal archiviert werden. Dies betrifft insbesondere Tonträger und audiovisuelle Medien. Während Nutzungsdigitalisate traditionell auf Laufwerken und/oder im Bildspeicher von FAUST vorliegen und als „Digitale Dokumente“ bzw. in Bildfeldern mit dem FAUST-Objekt des Originals verknüpft sind, stellte sich die Frage, wie mit in DiPS.kommunal archivierten digitalen Repräsentationen bei der Erschließung verfahren werden soll. Zum einen sollten die für alle Repräsentationen gleichermaßen gültigen inhaltlichen Erschließungsinformationen konsistent bleiben. Zum anderen sollten die jeweils spezifischen Angaben zum Trägermaterial bzw. zu den technischen Eigenschaften der digitalen Repräsentation, soweit sie für die Erschließung relevant sind, eindeutig der jeweiligen Ausprägung zugeordnet werden können.

Schließlich sollte neben diesen für die alltägliche Arbeit relevanten Anforderungen und Fragestellungen wie bei jedem Fachverfahren auch eine Exit-Strategie mitbedacht werden.

Die Umsetzungsmöglichkeiten für die hier in Auswahl beschriebenen Anforderungen wurden sowohl im Hinblick auf eine Weiterentwicklung der bestehenden Verzeichnungsdatenbank als auch mit Blick auf die Einführung einer zweiten Verzeichnungsdatenbank für digitales Archivgut in DiPS.kommunal durchdacht und einander gegenübergestellt. Der Vergleich ergab, dass bei beiden Varianten an der einen oder anderen Stelle Schwierigkeiten zu erwarten sind. Es zeigte sich aber auch, dass die Nutzung zweier Verzeichnungsdatenbanken sinnvoll ist und die Umsetzung gelingen kann, wie am folgenden Lösungsvorschlag erkennbar wird.

3 Der Lösungsvorschlag und seine aktuelle Umsetzung

Für die Erschließung von Archivgut, das in DiPS.kommunal archiviert wird, wurde die Datenbankstruktur in FAUST um eine weitere Verzeichnungsdatenbank ergänzt, die als Arbeitstitel bzw. archivintern als DiPS-DB bezeichnet wird. Für die Erschließung des Archivguts sind nun drei Datenbanken von zentraler Bedeutung: die schon bestehenden Datenbanken „Beständeübersicht“ und „Beständedatenbank“ sowie die neu angelegte „DiPS-DB“ (Abb. 1).⁷

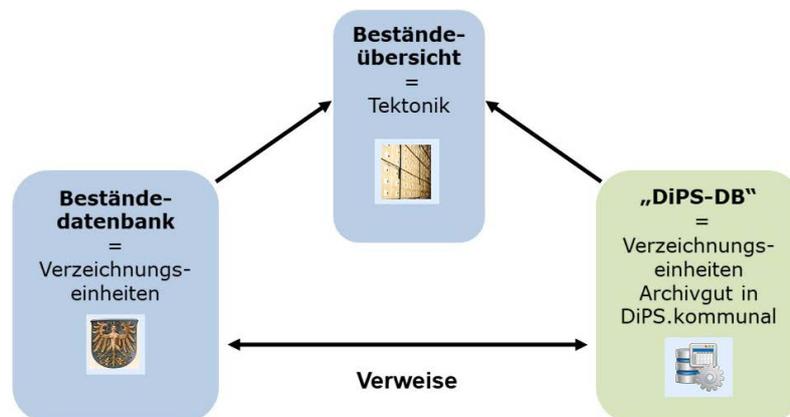


Abb. 1: Datenbanken zur Erschließung von Archivgut im Stadtarchiv Nürnberg. Grafik: Annette Birkenholz, Stadtarchiv Nürnberg.

Die DiPS-DB ist wie die Beständedatenbank über ein Referenzfeld mit der Beständeübersicht verknüpft. Die Beständeübersicht selbst wurde nur leicht angepasst, wobei die Überlegungen, welche Angaben zu digitalem Archivgut notwendig und sinnvoll sein könnten, noch nicht abgeschlossen sind. Dies betrifft beispielsweise eine gesonderte Erfassung von Umfangsangaben⁸ sowie Vormerkungen für Bestandserhaltungsmaßnahmen. Prinzipiell werden die Bestände des Stadtarchivs Nürnberg nach Provenienz gebildet und das digitale Archivgut wird den gleichen Beständen zugeordnet wie das

⁷ Die Datenbanken zur Tiefenerschließung, insbesondere die Datenbank zur Erfassung und Recherche von Personendaten etwa aus der standesamtlichen Überlieferung, spielen für die zunächst vorgenommene allgemeine Konzeption der Erschließung von digitalem Archivgut keine Rolle. Nicht dargestellt sind außerdem verbliebene Altdatenbanken, für die eine Integration in die Beständedatenbank beabsichtigt ist (Filmdatenbank).

⁸ Bei der Diskussion im Rahmen des AUdS-Camps wurde u. a. auf die Schwierigkeiten bei der Erfassung von Größen verwiesen, da diese etwa durch Formatmigrationen Veränderungen unterliegen. Im Stadtarchiv Nürnberg wurde festgestellt, dass allein die Festlegung einer einheitlichen, gut lesbaren Maßeinheit problematisch ist, da die Spanne je nach Bestand und Archivaliengattung von MB bis TB reicht.

analoge Archivgut. Beispielsweise werden die ab 1999/2000 vorliegenden Dateien mit den Tonaufnahmen der Stadtrats-sitzungen, eines der ersten Übernahmeprojekte des Stadtarchivs Nürnberg, dem Bestand „C 85/III – Bürgermeisteramt / Stadtratsprotokolle ab 1959“ zugeordnet, der bislang die auf Papier festgehaltenen Sitzungsniederschriften dieser Zeit umfasst. Eine eventuelle Bildung eigener Teilbestände, ähnlich wie auch bisher schon Abgrenzungen vorgenommen wurden, ist damit aber nicht ausgeschlossen.

Um ein einheitliches Nachweissystem zu erreichen, werden für die Bestandssignatur und die Klassifikation Zentralthesauri verwendet. Gleiches gilt für den Sperrgründethesaurus, der rechtliche, vertragliche oder dem Erhaltungszu-stand geschuldete Begründungen für Schutzfristen und andere Beschränkungen der Vorlage von Archivalien enthält. Durch die Nutzung von Zentralthesauri werden die in der Beständedatenbank gepflegten Thesauri auch in der DiPS-DB verwendet (Abb. 2).

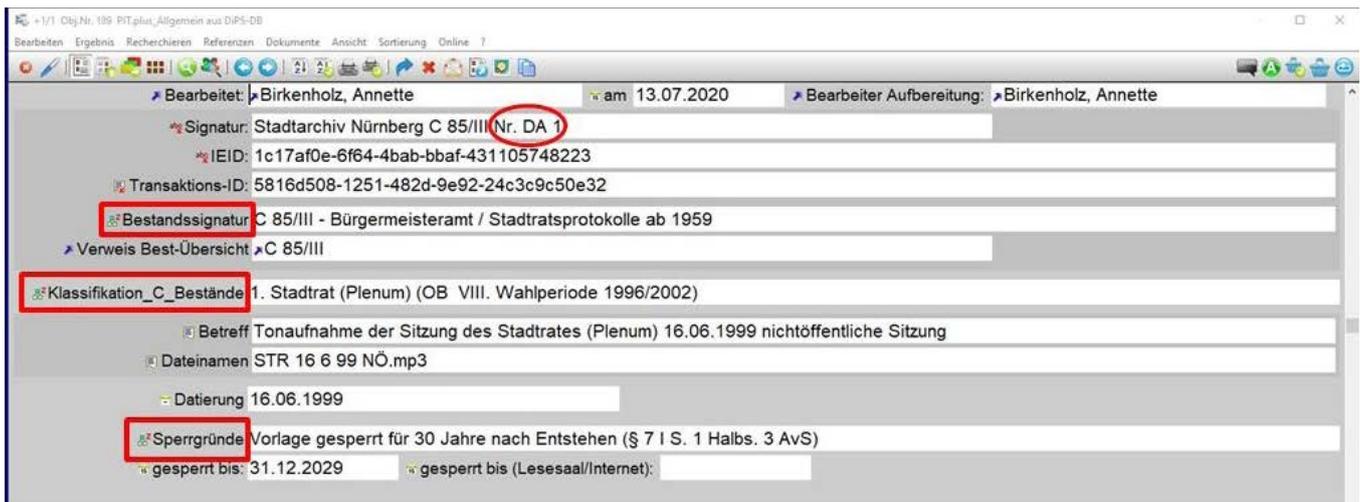


Abb. 2: Erschließungsmaske in der DiPS-DB, hervorgehoben sind die Signatur und die Nutzung von Zentralthesauri. Screenshot/Grafik: Annette Birkenholz, Stadtarchiv Nürnberg.

Hinsichtlich der Signaturen wurde entschieden, dass in der DiPS-DB eine eigene fortlaufende Zählung der Archivalieneinheiten, für jeden (Teil-)Bestand beginnend bei 1, eingeführt werden soll. Damit sollen doppelte Signaturen, Springnummern und der Aufwand für den Abgleich der Nummerierung zwischen Beständedatenbank und DiPS-DB vermieden werden. Zur Unterscheidung von den Signaturen in der Beständedatenbank wird das Kürzel „DA“ für digitales Archiv der fortlaufenden Nummer vorangestellt. Zugleich ist dies ein Hinweis darauf, dass es sich um digitales Archivgut in DiPS.kommunal handelt, was möglicherweise einen Vorteil für Bestellprozesse darstellen kann, die jedoch erst noch konzipiert werden.

Eine Einschränkung für die Umsetzung des gemeinsamen Nachweissystems über zwei Verzeichnungsdatenbanken hinweg musste jedoch in Kauf genommen werden: Anders als bei den Thesauri sind die Indexlisten, die für die Verschlagwortung von Personen- und Ortsnamen sowie Sachbegriffen verwendet werden, immer nur in der jeweiligen Datenbank verfügbar. Die eigenständigen Indices von Beständedatenbank und DiPS-DB werden jedoch als hinnehmbar betrachtet. Denn zum einen schlägt sich in den Indices der Beständedatenbank trotz Konventionen ein recht kreativer Umgang nieder, sodass ein regelkonformer Neuaufbau der Indices in der DiPS-DB auch als vorteilhaft angesehen wird. Zum anderen ist mit Einschränkungen immer noch eine übergreifende Suche in den Indices möglich, wie noch im Zusammenhang mit der Recherche gezeigt wird.

Für die Verzeichnung von hybridem Archivgut und Repräsentationen wurden gegenseitige Verweise zwischen Beständedatenbank und DiPS-DB als Lösung gewählt. Die Verweisfelder werden in der Beständedatenbank als „Verweis_DiPS-DB“ und in der DiPS-DB als „Verweis_BDB“ bezeichnet (Abb. 3). Für die Objektarten in der neu eingerichteten DiPS-DB ist ein Auswahlfeld vorgesehen, um zu kennzeichnen, dass es sich um hybrides Archivgut handelt. In der Beständedatenbank wird hingegen weiterhin im Bemerkungsfeld ein ausformulierter Hinweis auf zugehöriges digitales Archivgut eingetragen.

	*Signatur: Stadtarchiv Nürnberg E 6/1283 (Nr. 40)
	*Bestandssignatur: E 6/1283 - Hörervertretung (HV) beim Bildungszentrum Nürnberg (OB E 6 - Vereinsarchive)
	➤ Verweis Best-Übersicht ➤ E 6/1283
	*Klassifikation_E_F_Bestände Hörervertreterversammlung (HVV) (OB E 6/1283 Hörervertretung (HV) beim Bildungszentrum Nürnberg)
	*Bestellnummer 40
	Betreff "hv aktuell" - Kurzberichte über die Hörervertreterversammlung
	Umfang/Beschreibung: ca. 80 Blatt
	Bemerkungen: 2 Disketten wurden unter den Signaturen Stadtarchiv Nürnberg E 6/1283 Nr. DA 1 und DA 2 digital archiviert.
	Datierung 1986 - 1998
	➤ Verweis_DiPS-DB ➤ Stadtarchiv Nürnberg E 6/1283 Nr. DA 1
	➤ Verweis_DiPS-DB ➤ Stadtarchiv Nürnberg E 6/1283 Nr. DA 2

	*Signatur: Stadtarchiv Nürnberg E 6/1283 Nr. DA 1
	*IEID: 960ab55e-7dfb-4e90-ba28-eeed03d9a4d9
	*Transaktions-ID: 78578faf-f7de-4ac9-86eb-7904ea507048
	*Bestandssignatur E 6/1283 - Hörervertretung (HV) beim Bildungszentrum Nürnberg
	➤ Verweis Best-Übersicht ➤ E 6/1283
	Betreff "hv aktuell" - Kurzberichte über die Hörervertreterversammlung, Ausgabe Nr. 30 2/99
	Dateinamen Hvakt30 - RTF.rtf
	Datierung 1999
	<input checked="" type="checkbox"/> Hybrides Archivgut ja, siehe Verweis_BDB
	➤ Verweis_BDB ➤ Stadtarchiv Nürnberg E 6/1283 Nr. 40

Abb. 3: Verzeichnung hybrider Archivalien in der Beständedatenbank (oben) und der DiPS-DB (unten). Screenshot/Grafik: Annette Birkenholz, Stadtarchiv Nürnberg.

Die Verwendung von zwei getrennten Datenbanken mit zwei unterschiedlichen Signatursystemen hat dazu geführt, dass der analoge und der digitale Bestandteil von inhaltlich zusammengehörigen Archivalien eine jeweils eigene Signatur mit unterschiedlicher Bestellnummer erhalten.⁹ Die gegenseitigen Verweise sind also essenziell für die Kenntnis der Zusammengehörigkeit.

Hybride Archivalien liegen derzeit dergestalt vor, dass einem Aktenordner o. ä. Datenträger beigelegt sind. Aber auch auf den umgekehrten Fall, dass zu einer führenden elektronischen Akte noch Papieranteile vorhanden sind, lässt sich diese Lösung nach den derzeitigen theoretischen Überlegungen anwenden.

Bei der Verzeichnung von Sicherungsdigitalisaten, die als weitere Repräsentation in DiPS.kommunal archiviert sind, werden ebenfalls gegenseitige Referenzfelder sowie ein Auswahlfeld in der DiPS-DB und das Bemerkungsfeld in der Beständedatenbank genutzt (Abb. 4). Das Auswahlfeld in der DiPS-DB ist auch für die Unterscheidung zwischen Ersatz- und Sicherungsdigitalisaten sowie genuin digitalem Archivgut erforderlich.

⁹ In einer Diskussion wurde angemerkt, dass dieses Vorgehen mit einer Einzelblattverzeichnung verglichen werden kann.

The image contains two screenshots of archival records. The top screenshot is from the Beständedatenbank and shows the following fields: Signatur: Stadtarchiv Nürnberg E 9/659 Nr. 54; Bestandssignatur: E 9/659 - Spätlese Verlag Nürnberg (OB E 9 - Firmen- und Wirtschaftsarchive); Verweis Best-Übersicht: E 9/659; Bestellnummer: 54; Betreff: Audiokassette "Fränkische Geschichten" mit der Lesung von Wolf Peter Schnetz; Umfang/Beschreibung: 1 Audiokassette, 1 Sicherungs-CD; Bemerkungen (Intern): Kassette wurde auf CD gebrannt; Bemerkungen: Digitalisat der Audio-Kassette in DiPS.kommunal (E 9/659 Nr. DA 3); Datierung: 2004; Verweis_DiPS-DB: Stadtarchiv Nürnberg E 9/659 Nr. DA 3. The bottom screenshot is from the DiPS-DB and shows: Signatur: Stadtarchiv Nürnberg E 9/659 Nr. DA 3; IEID: 2f8269fc-1916-4cfe-9001-45352838f08c; Transaktions-ID: 1b32531b-b536-4cb0-bbaf-a0ab99ce9115; Bestandssignatur: E 9/659 - Spätlese Verlag Nürnberg; Verweis Best-Übersicht: E 9/659; Betreff: Audioaufnahme "Fränkische Geschichten" mit der Lesung von Wolf Peter Schnetz, L. Vogel und Matthias Kröner; Dateinamen: Fränkische Geschichten_19.09.04.wav; Datierung: 2004; Digitalisat: Sicherungsdigitalisat; Weitere Speicherorte: ja, siehe Verweis_BDB; Verweis_BDB: Stadtarchiv Nürnberg E 9/659 Nr. 54.

Abb. 4: Verzeichnung von Repräsentationen in der Beständedatenbank (oben) und der DiPS-DB (unten). Screenshot/Grafik: Annette Birkenholz, Stadtarchiv Nürnberg.

Eine Gesamtbetrachtung der Umsetzungsmöglichkeiten des Repräsentationenmodells vor dem Hintergrund der Gestaltung der FAUST-Nutzung im Stadtarchiv Nürnberg und im Zusammenspiel mit der Nutzung von DiPS.kommunal als Langzeitarchivierungssystem ist noch nicht erfolgt.¹⁰ Jedoch ist eine Erweiterung der DiPS-DB um einen Hinweis auf zusätzliche in DiPS.kommunal vorliegende Nutzungsrepräsentationen geplant.

Als weiterem Aspekt, der für die Umsetzung und das Gelingen der Zwei-Verzeichnungsdatenbanken-Lösung relevant ist, wird zum Schluss auf die Recherche eingegangen. Die Recherche erfolgt am komfortabelsten über die Suchmaske „Suche in allen Datenbanken“, was auch den Recherchegewohnheiten vieler Nutzer entsprechen dürfte. Mit der „Suche in allen Datenbanken“ kann in der Beständedatenbank, der DiPS-DB und weiteren dafür eingerichteten Datenbanken gleichzeitig recherchiert werden. Mittels der Eingabe von Suchbegriffen können auch die Indices von Beständedatenbank und DiPS-DB gleichzeitig durchsucht werden. Das Scrollen in den Indexlisten ist allerdings nur in der jeweiligen Datenbank möglich, sodass bei dieser Recherchetechnik beide Datenbanken jeweils für sich durchsucht werden müssen. Ein ähnlicher Nachteil ergibt sich bei der Recherche über die Thesauri mit der Bestandssignatur bzw. den Klassifikationen. Bei der Erfassung ist zwar sichergestellt, dass der identische Thesaurus benutzt wird. Bei der Recherche müssen die Thesauri aber in der jeweiligen Datenbank aufgerufen werden. Die Darstellung eines Gesamtergebnisses kann derzeit nur über die Zusammenführung der Ergebnisse aus beiden Datenbanken in einem Ergebniskorb erzielt werden. Wie nachteilig die genannten Einschränkungen für die Recherche empfunden werden, hängt auch mit den bevorzugten Recherchestrategien zusammen. Zukünftige Verbesserungen am Recherchesystem von FAUST sind zudem möglich.

10 Zum Repräsentationenmodell vgl. Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder (KLA), Das Repräsentationenmodell als Konzept für die archivische Arbeit. Thesenpapier, [September 2015], online unter: https://www.bundesarchiv.de/DE/Content/Downloads/KLA/repraesentationsmodell.pdf?__blob=publicationFile (abgerufen am: 14.07.2023).

4 Fazit

Dem Lösungsvorschlag liegt eine Abwägung des Stadtarchivs Nürnberg zugrunde, bei der der Aufwand für die Überarbeitung der überkommenen Strukturen der Beständedatenbank sowie die erwarteten Probleme mit der Systemperformance gegen die punktuellen Komforteinbußen bei der Recherche abgewogen wurden. Auch Bedenken, ob Verwirrung bei der Erschließung oder der Nutzung auftreten könnte, konnten entkräftet werden. Letztlich haben sie nicht gegen die Einrichtung einer zweiten Verzeichnungsdatenbank gesprochen. Gleiches gilt für die Frage, ob das neue Signatursystem langfristig zu Schwierigkeiten führen kann und ob ein Wechsel des Erschließungssystems durch die gewählte Lösung zusätzlich erschwert wird. Im Hinblick auf letzteres wurde jedoch festgestellt, dass auch an vielen anderen Stellen Referenzfelder in FAUST genutzt werden, sodass für den Erhalt dieser Informationen ohnehin Lösungen benötigt werden. Trotzdem sind Beobachtungen nötig, beispielsweise hinsichtlich der Auswirkungen auf digitales Archivgut, das (vorläufig) nicht in DiPS.kommunal archiviert wird. Dies betrifft insbesondere Ersatz- und Sicherungsdigitalisate, für die eine Archivierung in DiPS.kommunal zum jetzigen Zeitpunkt aufgrund der Datenmenge nicht infrage kommt und die vorerst nur mittels einer WORM-Speicherlösung gesichert und als „Digitale Dokumente“ in der Beständedatenbank verlinkt werden sollen.

Die Konzeption der DiPS-DB ist aber noch nicht beendet. Insbesondere sind noch die Erschließungsmasken zu konzipieren. Diese werden künftig – anders als bei der derzeitigen Mischnutzung, bei der die Masken sowohl nach Archivaliengattungen als auch nach Beständen organisiert sind – nur noch nach Archivaliengattungen eingerichtet werden. Nur in Einzelfällen, etwa für Zeitzeugeninterviews oder aufgrund besonderer datenschutzrechtlicher Anforderungen, sollen spezielle zusätzliche Objektarten eingerichtet werden. Auch praktische Schritte sind noch zu meistern, beispielsweise die Übertragung der bereits in der Beständedatenbank verzeichneten digitalen Fotos oder die Freischaltung der DiPS-DB für die Online-Recherche im iServer. Und schließlich stehen konzeptionelle Arbeiten für weitere FAUST-Datenbanken bevor, insbesondere für die Erfassung elektronischer Publikationen in der Bibliotheksdatenbank.

Gewerbekatasterdaten als Pilotprojekt – Die digitale Langzeitarchivierung beim Kommunalen Rechenzentrum Niederrhein (KRZN) mit startext SORI

Dr. Michael Habersack, Kreisarchiv Viersen, Fabian Näser, Startext GmbH

1 Einleitung

Das Kommunale Rechenzentrum Niederrhein (KRZN) ist das gemeinsame Rechenzentrum von insgesamt 45 kommunalen Gebietskörperschaften: den Kreisen Kleve, Viersen und Wesel jeweils einschließlich aller kreisangehörigen Gemeinden, dem Kreis Mettmann sowie den kreisfreien Städten Bottrop und Krefeld. Beim KRZN als ihrem gemeinsamen Dienstleister arbeiten die Mitgliedskörperschaften in zahlreichen Arbeitskreisen zusammen, um Anforderungen und Bedarfe gemeinsam zu formulieren, zu priorisieren und gemeinsam mit dem KRZN Lösungen zu finden, die das Rechenzentrum wiederum umsetzt und bereitstellt. 2012 wurde in diesem Kontext aus einem früheren Facharbeitskreis ein Unterarbeitskreis (UAK) für die Bearbeitungen aller Themen mit Bezug auf die digitale Langzeitarchivierung gegründet. Seine Zielsetzung war zunächst der Aufbau eines Verfahrenskatasters, die archivische Bewertung von eingesetzten Fachverfahren insgesamt und die Erarbeitung von Bewertungsmodellen für Verfahren mit archivwürdigen Daten in Erwartung der Produktivsetzung des Digitalen Archivs NRW (DA NRW)¹.

Nach der Erhebung der vorliegenden Situation im Verbandsgebiet folgte Anfang 2018 eine Erweiterung der Betrachtungsweise. Mit Blick auf den seit 2009 andauernden Entwicklungsprozess des DA NRW begann der Unterarbeitskreis eine Sondierung des Marktes der kommerziellen und öffentlichen Lösungen für die digitale Langzeitarchivierung einschließlich des DA NRW. Als Ergebnis seiner Analyse legte er dem Koordinierungskreis als gemeinsames Entscheidungsgremium der Mitgliedskörperschaften die Empfehlung zum Aufbau eines Testsystems im KRZN mit der Softwarelösung SORI der startext GmbH² vor. Innerhalb des für 2020 und 2021 angesetzten Pilotprojektes sollte die Funktionsweise der Softwarelösung für ein Fachverfahren erarbeitet und für den Produktivbetrieb vorbereitet werden. Im Zuge des Pilotprojektes wurde das bereits bewertete Gewerbekataster-Verfahren als Testverfahren ausgewählt.

2 Ausgangslage im KRZN

Durch die Größe des Verbandsgebietes und die Zahl der kommunalen Gebietskörperschaften werden sehr viele und sehr unterschiedliche Fachverfahren im Kommunalen Rechenzentrum Niederrhein betrieben. In vielen Fällen ist eine Dokumentenverwaltungsfunktionalität nicht oder nicht im Umfang der Anforderungen der nutzenden Fachverwaltungen vorhanden. Das KRZN hat daher eine Vielzahl der bei ihm betriebenen Verfahren an das eRegistrierungssystem Saperion angebunden, das für das Dokumentenmanagement genutzt wird. Für das DMS bzw. Registrierungssystem Saperion hat das KRZN eigene Entwicklungslizenzen und Entwickler. Durch diese Konstellation hat die Kombination aus vielen Fachverfahren und einem eRegistrierungssystem mehrere Vorteile: Der Datenstrom der einzelnen Fachverfahren in das eRegistrierungssystem kann durch das KRZN selbst modelliert und gesteuert werden; das gilt sowohl für den Datenstrom nach Saperion als auch für die Daten, die aus Saperion an ein digitales Langzeitarchiv abgegeben werden. Für die Modellierung maßgeblich sind die fachlichen Anforderungen des UAK; die Entwickler passen bei Bedarf die Saperion-Schnittstellen in beiden Systemen an.

Neben der technischen Umsetzung der Aussonderung aus den Fachverfahren durch das KRZN war die Hauptaufgabe des Unterarbeitskreises für das Pilotprojekt daher die fachliche Konzeption der Aussonderungen. Die Möglichkeit, die Exporte aus Saperion anpassen zu können, stellte sich im Pilotprojekt als ein großer Gewinn für den gesamten Systemaufbau dar, da durch die erweiterten Funktionalitäten für die Dokumentenverwaltung in Saperion die Möglichkeit besteht, die Daten aus den Fachverfahren weiter durch eDokumente oder Scans anzureichern. Auch hier ist der Vorteil einer zentralen Dokumentenverwaltungssoftware hervorzuheben. Auf der Grundlage der vorliegenden Systemlandschaft entschied sich der UAK dafür, Daten nicht hybrid aus dem Produktivsystem einerseits und ergänzend aus Saperion andererseits zu übernehmen, sondern gebündelt aus Saperion heraus die Daten an das Langzeitarchiv zu übergeben. Hierfür sprachen zwei zentrale Aspekte: In Saperion werden die Daten in den Formaten PDF und XML verwahrt. Beide bieten eine sehr gute Voraussetzung für die Langzeitarchivierung. Außerdem ist eine gute Nachnutzbarkeit von Teilen der Aussonderungsschnittstelle zu erwarten, wodurch eine hohe Effizienz bei Aufwand und Kosten erreicht werden kann, zumal aufgrund der ausschließlich in Saperion vorhandenen, gescannten oder gespeicherten Daten eine Schnittstelle aus dem DMS ohnehin zwingend erforderlich ist. Derjenige Teil der Fachverfahren, der nicht an Saperion angebunden ist, muss später gesondert betrachtet werden.

1 <https://www.danrw.de/> (abgerufen am: 14.07.2023)

2 <https://www.startext.de/produkte/sori> (abgerufen am: 14.07.2023)

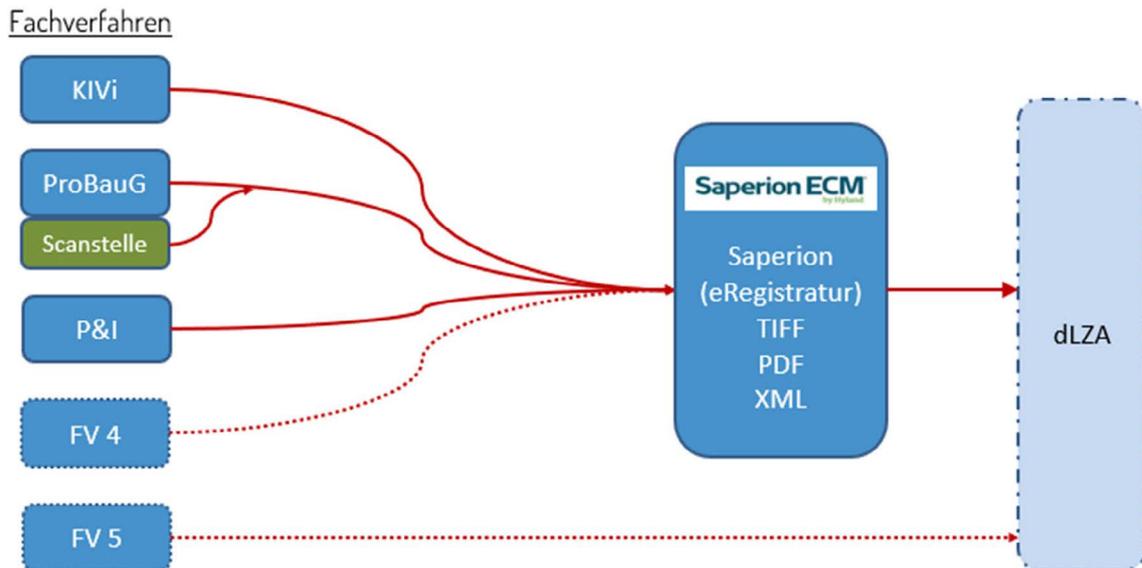


Abb. 1: Datenfluss im KRZN (Dr. Habersack, 2021)

Innerhalb des Fachverfahrens für die Gewerbesteuerdaten liegen in Saperion 16 anwenderübergreifend benutzte Standard-Dokumentarten inklusive der eingescannten Alt-Akten sowie bis zu 30 optional benutzte individuelle Dokumentarten vor. Der erste Schritt des Unterarbeitskreises in der Bearbeitung des Fachverfahrens lag in der grundsätzlichen Bewertung der Archivwürdigkeit der einzelnen Dokumentarten. Nach tiefergehender Analyse wurden – neben den individuellen – insgesamt sieben der 16 Dokumentarten für archivwürdig befunden:

- Anschreiben ausstehende Messbescheide
- DTA -Beleg
- GwSt-Bescheid
- Zinsbescheid
- Mitteilung Gewerbesteuermessbetrag
- GwSt-Berechnung
- Zinsberechnung

Als nicht archivwürdig wurden die folgenden Dokumentarten befunden:

- DTA-Beleg Verworfen
- DTA-Beleg Eingetragen
- GwSt-Bescheid Storno
- Zinsbescheid Storno
- Mitteilung Gewerbesteuermessbetrag Verworfen
- Mitteilung Gewerbesteuermessbetrag Eingetragen
- GwSt-Berechnung Storno
- Zinsberechnung Storno
- Sollliste

Im Pilotprojekt wurde früh entschieden, keine Registraturschnitte durchzuführen und dadurch Unterlagen aus offenen Vorgängen zu übernehmen, sondern nur Daten von abgeschlossenen Vorgängen auszusondern. Diese perspektivisch nicht in allen Fällen praktikable Vorgehensweise ist im Fall der Gewerbesteuerdaten möglich und bringt den Vorteil, dass keine redundanten Daten im Langzeitarchiv erzeugt werden. Weiter stellt sich die fachliche Frage nicht, ob bestehende AIPs im Langzeitarchiv nachträglich angereichert werden dürfen oder nicht.

Dadurch, dass es sich bei Saperion um ein DMS oder Registratursystem handelt, funktioniert hier vieles wie in der analogen Aktenwelt, was das Vorgehen für das Pilotprojekt vereinfachte. Hierzu gehört die Tatsache, dass die aus den Fachverfahren nach Saperion abgegebenen Daten unter einem digitalen „Aktendeckel“ gesammelt werden. Der „Aktendeckel“ selbst ist eine in Saperion erstellte Aggregation aus recherche- und verwaltungsrelevanten Daten aus den unter ihm zusammengeführten Daten und Dokumenten. Er stellt somit eine vorarchivische Erschließungsebene dar und schafft sinnvoll abgegrenzte Einheiten mit denormalisierten und dadurch strukturell wenig komplexen Daten.

Dadurch stehen für die folgende Übernahme in das dLZA nicht nur geeignete Datenpakete bereit, sondern auch die erschließungsrelevanten Metadaten für die verwendeten Archivinformationssysteme (AFIS). Dies spielte bei der späteren Entscheidung für die Struktur und den Umfang der Erschließungsmetadaten eine wichtige Rolle.

3 Die technische Aussonderung der Daten

Nachdem die fachliche Entscheidung über die archivwürdigen Dokumentarten getroffen worden war, bestand der nächste Schritt in der Überlegung, wie die Daten technisch aus Saperion ausgesondert werden sollen. Hier konnte der Vorteil genutzt werden, dass durch das KRZN die Aussonderungsschnittstelle von Saperion angepasst werden konnte. In einem intensiven Austauschprozess mit der startext GmbH entschieden sich UAK und KRZN, dass pro Ablieferung aus einem Fachverfahren eine XML-Datei erzeugt werden soll, in der mehrere Einträge (Dokumente) zu einer Gewerbeeinheit enthalten sind. Pro Eintrag können mehrere Dateien, zum Beispiel Scans der Originalakten, verknüpft werden.

4 Aufbau der Systemlandschaft im KRZN

Eine eigene Herausforderung stellte die kommunale Systemlandschaft im KRZN-Verbandsgebiet dar. Für die Recherche und für den Aufruf der Nutzungskopien (DIP) für die archivinterne Nutzung ist die Verwendung eines Archivinformationssystems (AFIS) hilfreich, da hier Informationen aus dem Langzeitarchiv im Kontext der gesamten Erschließungsinformationen eines Archivs recherchiert werden können. Vor allem aber war es für den UAK schon bei der Markterkundung ein wichtiges Auswahlkriterium, dass analoges und digitales Archivgut in einem System erschlossen und recherchiert werden kann. Konkret ging es um die Integrierbarkeit der drei im Verbandsgebiet verwendeten AFIS ACTApro der startext GmbH, AUGIAS der Firma AUGIAS-Data GmbH³ sowie FAUST der Firma LAND Software-Entwicklung⁴. Aufgrund dieser Ausgangslage wurde zu Beginn des Projekts ausführlich über die unterschiedlichen Möglichkeiten des Systemaufbaus auf Basis der bestehenden Strukturen diskutiert. Direkte Anbindungen aller verwendeten AFIS an SORI wurden mit Blick auf zukünftige Anpassungen an weitere Übernahmeprojekte und die zunächst unabhängige Entwicklung aller Softwares als aufwendig und potenziell problembehaftet eingeschätzt. Der UAK entschied sich daher für eine Lösung mit einem Ingestclient, der nicht nur bei der Datenübernahme verwendet wird, sondern die erschließungsrelevanten Metadaten zugleich an das verwendete AFIS vermittelt, also ähnlich einem Resolver zwischen den Systemen übersetzt.

Als Ingestclient bot sich hierbei das startext-AFIS ACTApro Desk⁵ an, weil es die gestellten Anforderungen bereits zum großen Teil erfüllte und „eh da“ war: Es kann sowohl digitale Objekte importieren als auch diese zur weiteren Verarbeitung an die dLZA-Software SORI übergeben, und es kann als AFIS Erschließungsinformationen in der Form als XML zur Verfügung stellen, die andere AFIS zur richtigen und vollständigen Darstellung der Erschließungsinformationen brauchen. Der Ingestclient kann die Saperion-Exporte aus den Fachverfahren gebündelt importieren und Verzeichnungsdatensätze erzeugen. Bereits an dieser Stelle besteht so die Möglichkeit, sowohl die „mitgebrachten“ Daten aus den Fachverfahren im Archivinformationssystem ACTApro Desk um weitere Erschließungsinformationen bei Bedarf anzureichern als auch nach den Daten zu recherchieren. Die Übergabe an das Langzeitarchiv erfolgt dabei vollautomatisiert aus ACTApro Desk heraus: Dabei werden die Verzeichnungsdaten inklusive angehangener Dateien an das Langzeitarchiv übergeben. Alle nach dem OAIS-Referenzmodell notwendigen Bearbeitungsschritte werden während des Ingests vollautomatisiert durchgeführt. Nach Rückmeldung der IDs zu den einzelnen Archivpaketen in die Verzeichnungsdaten ist zudem ein Export der Verzeichnungsdaten im EAD-Format aus dem Ingestclient ACTApro Desk heraus vorgesehen. Die EAD-XML können dann in AUGIAS und FAUST importiert werden. Der Zugriff auf die Nutzungskopien im Langzeitarchiv kann direkt aus AUGIAS und Faust erfolgen.

3 <https://www.augias.de/> (abgerufen am: 14.07.2023)

4 <https://www.land-software.de/> (abgerufen am: 14.07.2023)

5 <https://www.startext.de/produkte/actapro> (abgerufen am: 14.07.2023)

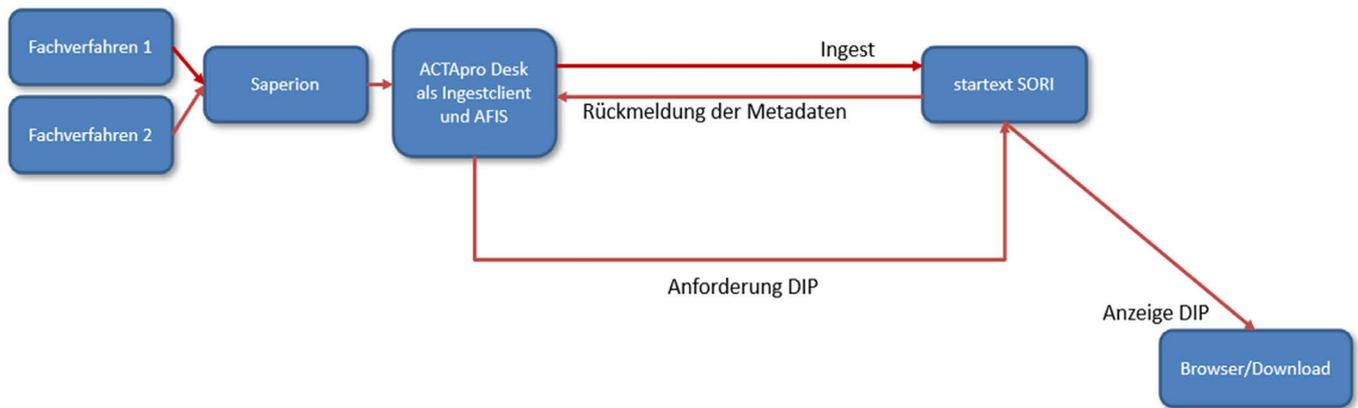


Abb. 2: Systemaufbau im KRZN (Näser, 2021)

5 Die Zielstruktur der Daten im AFIS

Im weiteren Verlauf des Pilotprojektes wurden die Gewerbesteuerdaten im Rahmen einer Testphase zuerst konzeptionell betrachtet und daraufhin die Übernahme nach ACTApro Desk programmiert. Um die Zielstruktur in ACTApro Desk zu definieren, war eine genaue fachliche Absprache der gewünschten Datenstruktur innerhalb des Unterarbeitskreises notwendig. Hierbei lag ein Augenmerk auf der vorhandenen Systemlandschaft: Die Daten mussten nicht nur im Ingestclient und im AFIS ACTApro Desk, sondern auch in den anderen Archivinformationssystemen angezeigt und dort auch recherchiert werden können. Als ersten Ansatz wurde eine 1:1 Übernahme der Dokumente aus Saperion diskutiert. Diese Variante hätte zur Folge gehabt, dass zu jedem Dokument und seinem Metadatensatz ein Verzeichnungsdatensatz mit entsprechend angepasster Verzeichnungsmaske in ACTApro Desk automatisch erstellt worden wäre. Der Vorteil dieses Ansatzes hätte in der sehr hohen Erschließungstiefe und Recherchierbarkeit gelegen, entsprechend einer analogen Einzelblattverzeichnung. Diese Variante wurde in einer ersten Version direkt in ACTApro Desk umgesetzt, um vonseiten des Unterarbeitskreises den kompletten Workflow vom Import bis zur Ansicht und Recherche überprüfen zu können. Der Ansatz wurde aus mehreren Gründen nach einer Testphase verworfen. Ausschlaggebend waren zwei Aspekte: Erstens hätte der Ansatz perspektivisch zu einer enormen Zahl von Verzeichnungseinheiten und damit auch AIPs geführt und jedes AFIS potenziell mit Metadaten von für sich oft wenig aussagekräftigen Einzelunterlagen „zugemüllt“. Der UAK befürchtete den Suchmaschinen-Effekt, dass man vor lauter Treffern das Gesuchte nicht mehr findet. Zweitens erlaubte der Ansatz keine Ordnung nach Betrieben. Die Reihenfolge der automatisiert angelegten Verzeichnungsdatensätze richtete sich nach dem Erstelldatum in Saperion, die Herstellung des Zusammenhangs der einzelnen Dokumente wäre nur mit weiterem Aufwand erreicht worden. Eine frühe Verknüpfung von Datensätzen in ACTApro Desk als Ingestclient wäre zwar möglich gewesen. Der Transport über die Systemgrenze hinweg in die beiden anderen verwendeten AFIS AUGIAS und FAUST hätte aber zusätzlichen Anpassungsaufwand impliziert und wäre mit dem als Austauschformat angedachten EAD-Format nicht abzubilden gewesen. Eine späte Verknüpfung erst in den jeweils verwendeten AFIS hätte dagegen ebenfalls ständigen Nachbearbeitungsaufwand bedeutet. Von Beginn an war das Ziel des Pilotprojekts aber, einen auch produktiv gangbaren Weg für dieses Fachverfahren zu ermitteln, auch wenn dies das Verwerfen von einzelnen Umsetzungsmöglichkeiten bedeutete.

Im zweiten Ansatz wurden die Anzahl der Verzeichnungsdatensätze und die Übersichtlichkeit in den Fokus genommen. Als mögliches Schlüsseldatum wurde das Kassenzeichen identifiziert. Dieser Ansatz sah vor, alle Dokumente zu einem Kassenzeichen in einem Verzeichnungsdatensatz zu bündeln, zu jedem Dokument aber unterhalb der Ebene der Verzeichnungseinheit einen Vorgang zu erstellen. Das Kassenzeichen wurde als Identifizierungsmerkmal eines Gewerbes gewählt, da dieses auch nach einem Namenswechsel des Gewerbebetriebs unverändert bleibt. Die Verwendung der Vorgangsebene bot den potenziellen Vorteil, für die Dokumente weiterhin zusätzlich zur Verzeichnungseinheit eine Einzelblattverzeichnung durchführen zu können. Zum Zeitpunkt der Testphase des Piloten war an dieser Stelle jedoch nicht klar, wie diese Struktur sinnvoll in AUGIAS und FAUST umgesetzt werden sollte. Weiterhin war dieser Ansatz in der technischen Umsetzung verhältnismäßig aufwendig sowie fachlich nicht für alle weiteren Fachverfahren nutzbar. Der Ansatz wurde daher ebenfalls verworfen und eine dritte Variante zwischen dem Unterarbeitskreis und der startext GmbH diskutiert.

In der dritten und finalen Variante wurde die Idee aus dem zweiten Anlauf aufgegriffen, eine Verzeichnungseinheit pro Gewerbebetrieb aus Saperion zu erstellen. Durch den Austausch im Projekt hatte sich herausgestellt, dass in Saperion nicht nur die Dokumente zu einem Gewerbebetrieb vorgehalten werden, sondern dass aus den nach Saperion übernommenen Daten auch verbindende, recherchierbare Metadaten aggregiert werden und dass der Export dieses

„Aktendeckels“ möglich ist. Der „Aktendeckel“ kann als Grundlage der Bildung der Verzeichnungseinheiten genutzt werden. Die Verzeichnungsmaske der Verzeichnungseinheit in ACTApro Desk wird dabei immer durch die Metadaten des Schlussdokuments, des Gewerbesteuer-Bescheids („GwSt-Bescheid“), befüllt. Die Vorzüge dieser Variante liegen in der Übersichtlichkeit der Verzeichnung, der Nachnutzbarkeit der Struktur sowie im einfacheren Ausspielen der Daten in die Archivinformationssysteme AUGIAS und FAUST.

Eine Konsequenz dieser Struktur ist, dass durch die automatisierte Übernahme der Metadaten aus dem Schlussdokument in die Verzeichnungseinheit ältere Stände über die Volltextsuche in den älteren Dokumenten gesucht werden müssen. Nach Betrachtung aller Vor- und Nachteile dieser Variante wurde diese final vom UAK freigegeben und von der startext GmbH umgesetzt.

Nach der finalen Absprache der Struktur bestand der nächste Schritt in einer Analyse, wie die Verzeichnungsmasken in ACTApro Desk aussehen bzw. welche Felder aus den Fachverfahren hier direkt sichtbar sein sollen. Diese Felder sind als Konsequenz nicht nur über eine Volltextsuche, sondern auch über eine feldspezifische Recherche auffindbar. Letztendlich wurde beim Erstellen der Verzeichnungsmasken das Augenmerk darauf gelegt, welche Daten für Nutzende bei der Recherche relevant sind. Um die Laufzeitrecherche zu erleichtern, wird außerdem aus den Daten der Gewerbeanmeldung und dem Abschlussdatum automatisiert eine Laufzeit erstellt und in das entsprechende Feld eingetragen.

6 Aufbau des Ingestworkflows

Nachdem die grundlegenden Arbeiten für die Datenstruktur abgeschlossen waren, wurde der Ingestworkflow für SORI eingerichtet. SORI bietet bereits eine vorkonfigurierte Verarbeitungskette mit den notwendigen Schritten.

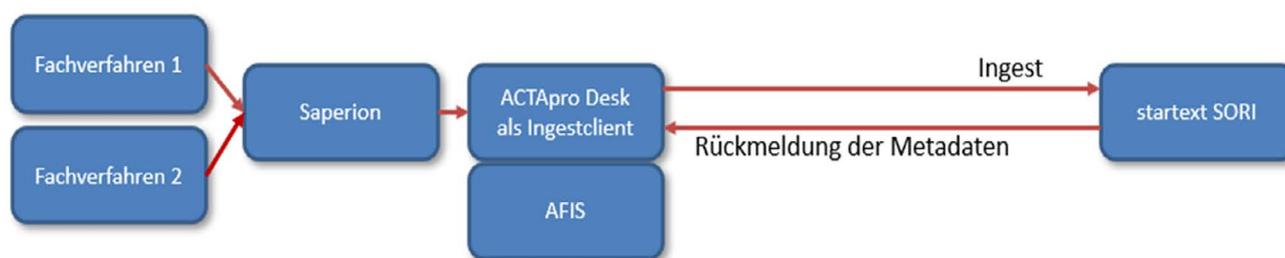


Abb. 3: Prozessworkflow im KRZN (Näser, 2021)

Der Ingest wird aus ACTApro Desk heraus gestartet: Die Verarbeitung der Archivpakete erfolgt automatisiert und asynchron serverseitig durch SORI. Nach erfolgreicher Bearbeitung aller vorkonfigurierten Ingestschritte erfolgt die Rückmeldung der erfolgreichen Bearbeitung an ACTApro Desk. Für die Verwaltung der einzelnen Ingestpakete bietet SORI eine eigene Weboberfläche, über die alle laufenden und abgeschlossenen Aktionen eingesehen werden können. Dabei unterscheidet SORI prozessbezogen immer zwischen Ingest-, Access- sowie Preservation-Prozessen. Über die Weboberfläche kann überprüft werden, in welchem Status sich ein Paket gerade befindet. Schlägt ein Prozessschritt fehl, zum Beispiel weil bestimmte Systemkomponenten nicht verfügbar sind, fällt der Prozess in den vorherigen Prozessschritt zurück. Von diesem Status aus kann der Ingest zu einem späteren Zeitpunkt jederzeit wieder neu gestartet werden. Jeder einzelne Prozess in SORI ist somit also transaktionssicher und kann bei einem Problem, auch bei systembedingten Problemen mit dem Server oder der Datenbank, jederzeit neu gestartet werden. Die Ergebnisse der einzelnen Prozessschritte werden auf der Weboberfläche zu jedem Prozess angezeigt. Neben der Anzeige auf der Weboberfläche werden alle Aktivitäten auch in einer PREMIS-XML-Datei protokolliert, welche im AIP mit abgespeichert wird. Auf diesem Weg wird auch jeder Neustart eines Prozesses bei einem Fehlerfall nachvollziehbar dokumentiert. Sobald der Ingest erfolgreich abgeschlossen ist, werden die Informationen zum Paket nach ACTApro Desk zurückgemeldet. Der Aufruf der Nutzungskopien kann dann direkt im Archivinformationssystem erfolgen, wobei jeder Zugriff als Access-Prozess protokolliert wird. Die Nutzungskopien werden in SORI bei Bedarf erstellt und nicht direkt als Teil des AIP im Langzeitarchiv abgelegt. Um hier die Zugriffsdauer auf die Nutzungskopien zu verringern, können bereits erstellte Nutzungskopien auf Wunsch serverseitig gecached werden, sodass die Nutzungskopien nur noch übertragen, nicht aber erst erstellt werden müssen.

Im Rahmen der DIP-Erstellung kann eine Umwandlung in bestimmte Darstellungsformate eingerichtet werden.

Dadurch dass alle Bearbeitungsschritte eines Paketes in SORI vollautomatisiert integriert sind, ist eine Einsichtnahme der Weboberfläche von SORI im Normalbetrieb nicht notwendig. Die einzelnen Bearbeitungsschritte sind im Rahmen der Konfiguration von SORI anpassbar, so können bei weiteren Anforderungen aus späteren Übernahmen weitere Schritte hinzugeschaltet werden, welche dabei automatisch mit in die Protokollierung auf der SORI Weboberfläche und

im PREMIS-XML mit aufgenommen werden. Auch die für die einzelnen Schritte eingesetzten Werkzeuge, etwa für Formatvalidierung oder Formatmigration, können ausgetauscht werden. Ein Beispiel für einen durch Änderung der Konfiguration hinzugefügten Bearbeitungsschritt ist das Ermitteln und Prüfen von fachlich festzulegenden signifikanten Eigenschaften vor und nach einer Formatmigration.

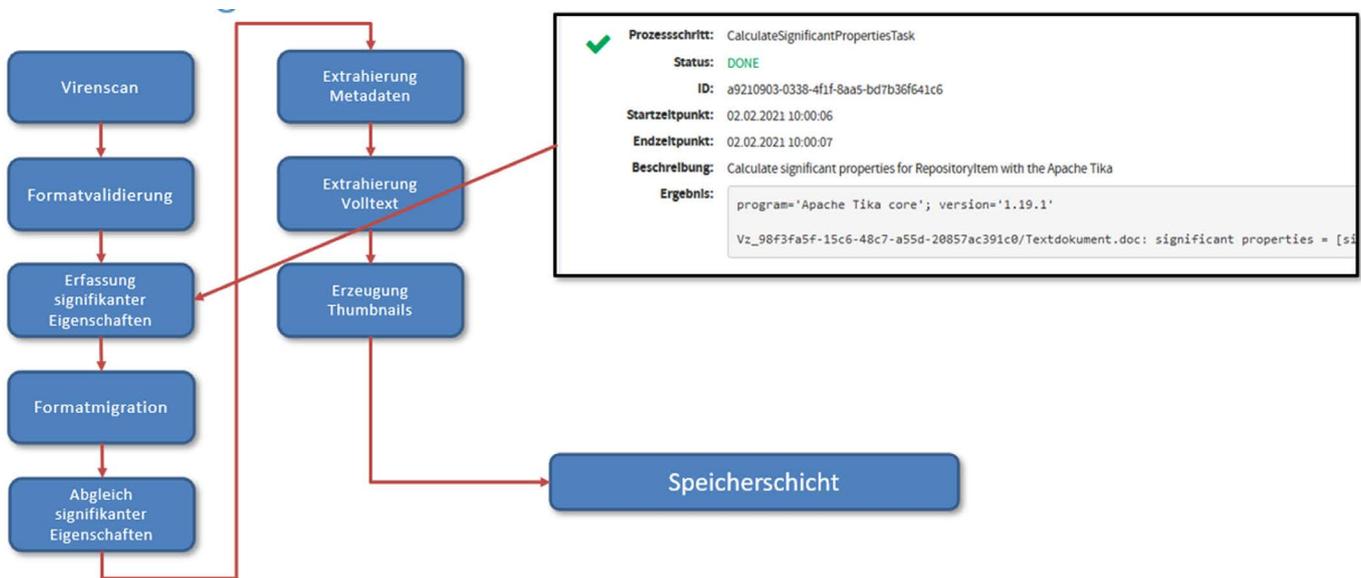


Abb. 4: Die konfigurierbare Verarbeitungskette in SORI (Näser, 2021)

Sobald alle Schritte erfolgreich durchgeführt wurden, wird das Paket auf der Speicherschicht von SORI abgespeichert. Die Informationen im Paket werden aufgeteilt in die Bereiche „content“ und „metadata“. Im Bereich „content“ werden die Primärdateien abgelegt. Diese werden jeweils in unterschiedlichen Repräsentationen gespeichert, sofern durch eine Preservation Action eine Formatmigration durchgeführt wird. Im Rahmen des Pilotprojekts wurde auch diskutiert, ob die Formatmigrationen für die PDFs aus Saperion bereits während des Ingests oder zu einem späteren Zeitpunkt durch eine Preservation Action durchgeführt werden sollen und ob beide Formate archiviert werden sollen, ob also neben dem archivtauglichen PDF/A zusätzlich das aus Saperion kommende PDF aufbewahrt werden soll. Durch die Formatmigration der übernommenen PDFs ist keinerlei Verlust zu erwarten. Der UAK entschied sich daher in diesem Fall gegen die Aufbewahrung beider Formate und die damit einhergehende Verdopplung des benötigten Speichervolumens. Ins Langzeitarchiv übernommen wird nur die formatmigrierte Datei. Im Bereich „metadata“ wird neben der PREMIS-Protokolldatei auch das XML der Verzeichnungseinheit aus ACTApro Desk heraus zusätzlich mit abgespeichert. Dies verbraucht sehr wenig Speicher, hat aber den Vorteil, dass bei einem Verlust der Erschließungsdatenbank durch die Informationen im Langzeitarchiv die Datenstruktur im Archivinformationssystem wiederhergestellt werden könnte. Die Archivpakete selbst werden aktuell physisch auf der Speicherschicht gespeichert. Die IDs der Pakete wird nach erfolgreichem Abspeichern an ACTApro Desk zurückgemeldet, sodass der Aufruf eines DIP direkt aus dem Archivinformationssystem heraus erfolgen kann. Die ID wird bei einem EAD-Export mit ausgegeben, sodass auch aus AUGIAS und FAUST heraus ein Aufruf der Nutzungskopien möglich ist.

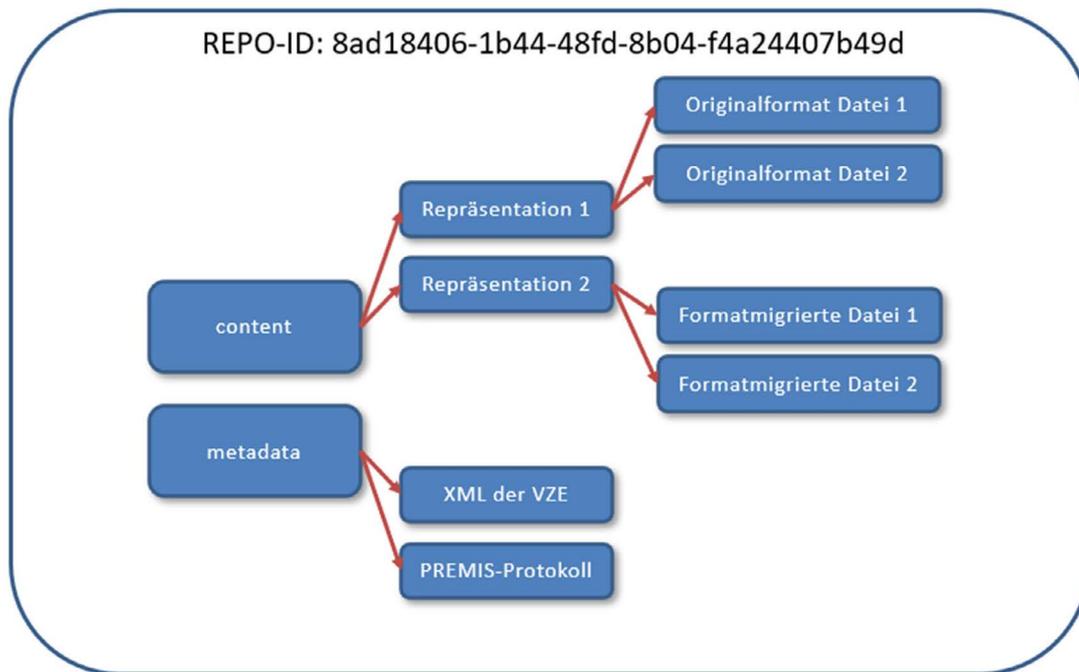


Abb. 5: Aufbau eines AIP in SORI (Näser, 2021)

7 Erreichter Stand und nächste Schritte

Mit der Fertigstellung der Workflows für das Gewerbesteuer-Fachverfahren konnte das Pilotprojekt erfolgreich zum Abschluss gebracht werden. Der UAK stellte dem Qualitätszirkel des KRZN die Testergebnisse vor und empfahl dem Koordinierungskreis den Aufbau des Produktivsystems eines digitalen Langzeitarchivs mit SORI. Der Koordinierungskreis beschloss den Aufbau im September 2021. Das Ausrollen von SORI für das Verbandsgebiet des KRZN ist noch für 2021 geplant.

Als nächstes steht die Übernahme von archivwürdigen Daten aus weiteren Fachverfahren an, wobei hier zunächst weitere, an Saperion angebundene Fachverfahren im Fokus stehen werden. In weiteren Schritten wird der UAK zusammen mit dem KRZN und der startext GmbH die Konzeption der Aussonderung und Übergabe von archivwürdigen Daten aus nicht an Saperion angebotenen Fachverfahren an SORI in Angriff nehmen. Viele der im Pilotprojekt gewonnenen Erkenntnisse in Bezug auf Importroutinen und die Darstellung der Daten in den verwendeten Archivinformationssystemen werden für die anstehenden Projekte weiterverwendet werden können. Hierzu gehört auch die Erkenntnis, dass aus elektronischem Archivgut nicht nur sehr leicht Erschließungsmetadaten generiert werden können, sondern dass auch die Grenzen zwischen beidem viel leichter verschwimmen als in der analogen Welt. Man sollte sich immer wieder der kritischen Frage stellen, welche Daten tatsächlich für die Erschließung gebraucht werden. Denn mit jeder verfahrensspezifischen Anreicherung steigt potenziell die Zahl der Erfassungsmasken und Pflegeaufwände. Bedeutete jede Aggregation von Daten in der analogen Welt einen Aufwand, den zu betreiben man als Gewinn sehen konnte, so scheint es ratsam, bei genuin elektronischen Daten konzeptionellen Aufwand für die Beschränkung dessen aufzuwenden, was in ein Archivinformationssystem gelangt.

8 Fazit

Die Übernahme von Fachverfahrensdaten im Pilotprojekt hat gut funktioniert. Das schließt die Möglichkeit des Aufrufs von DIPs für den internen Gebrauch im Archivinformationssystem bereits mit ein. Parallel zu den anstehenden Übernahmeprojekten für Fachverfahrensdaten wird es eine weitere Aufgabe sein, Nutzungskopien auch in einem Digitalen Lesesaal bereitstellen zu können. Die digitale Langzeitarchivierung bleibt damit auch in den Kommunalarchiven des Niederrheins und im KRZN als ihrem IT-Dienstleister eine große Aufgabe mit vielen offenen Fragen in den weiteren Projekten, aber der konkrete Einstieg in die Praxis des digitalen Archivierens ist mit substantiell greifbaren Ergebnissen in sehr überschaubarer Zeit gelungen und stimmt optimistisch für die Zukunft.

Archivierung von Daten aus Twitter: Das „Twitterarchiv“ über die Twitter-eigene Exportschnittstelle 2018–2021

Stephanie Kortyla, Sächsisches Staatsarchiv

1 Stellenwert sozialer Medien

Laut D21-Digital-Index 2020/2021 werden in Deutschland im Durchschnitt pro Kopf drei soziale Medien bzw. Netzwerke genutzt. Dies schließt aktive als auch passive Nutzung ein. Soziale Medien dienen der Rezeption, Information, Interaktion sowie Partizipation. Synonym wird auch von „Mitmach-Web“ gesprochen. Verschiedene Altersgruppen nutzen dabei unterschiedliche soziale Medien. Jüngere Nutzer, digital natives, sind derzeit vorrangig bei Instagram und Tiktok anzutreffen, die Generation X, also geboren zwischen 1960 und 1980, dagegen öfters bei Facebook. Gleichermaßen aktiv sind alle bei YouTube und Twitter.¹ Die ersten sozialen Netzwerke entstanden in den 1980er-Jahren, einige heutige soziale Medien sind inzwischen seit 20 Jahren verfügbar. Sie unterliegen stetem technologischem Wandel, übernehmen z. B. gegenseitig beliebte Funktionen.² Insgesamt machen sie inzwischen einen Großteil der Netzressourcen aus.

Soziale Medien werden nicht nur im privaten Umfeld genutzt, auch Wirtschaft, NGO und öffentlicher Dienst nutzen diese Kanäle, um bestimmte Zielgruppen zu erreichen.

2 Archivierung von Netzressourcen im Sächsischen Staatsarchiv

Weil Verwaltungshandeln somit auch über soziale Medien abgebildet wird, beschäftigt sich das Sächsische Staatsarchiv seit 2018 mit der Archivierung von Daten aus Twitter. Es ergänzt damit seine bisherige Überlieferung von Netzressourcen. Seit 2014 werden Webauftritte anbieterpflichtiger Stellen turnusmäßig gecrawlt. Eventbasierte Crawls werden z. B. bei Landtagswahlen zusätzlich durchgeführt. Aktuell wird das Themenportal zum Coronavirus³ regelmäßig gecrawlt.

3 State of the art – Twitter & Archive

Zwischen 2006 und 2017 hat die Library of Congress Daten aus Twitter flächendeckend archiviert, seit 2018 aus Kapazitätsgründen nur noch selektiv. Bis heute sind die archivierten Daten der Öffentlichkeit nicht zugänglich.⁴ Bisherige Ansätze zur Archivierung von Twitter im deutschsprachigen Raum⁵ sind zu finden bei Thomas Wildi (2019)⁶, Peter Worm (2020)⁷ und dem Hauptstaatsarchiv Wiesbaden (2020)⁸. Beim globalen Schnittstellen- bzw. Tool-Tracking zur Twitterarchivierung geriet die Web Science and Digital Libraries Research Group (@WebSciDL)⁹ an der Old Dominion University (US-VA) in den Fokus. Praktische Beispiele bzw. einsehbare Webarchive existieren für den stillgelegten Twitter-Account von Donald Trump.¹⁰ Seit Anfang 2020 ist das Dateiformat „tweet.js(on)“ offiziell in der

1 Initiative D21 e.V. (Hrsg.), D21-Digital-Index 2020/2021 – Jährliches Lagebild zur Digitalen Gesellschaft, Berlin 2021, online unter: https://initiated21.de/app/uploads/2021/02/d21-digital-index-2020_2021.pdf (abgerufen am: 14.07.2023). Vgl. auch ARD/ZDF-Forschungskommission (Hrsg.), Nutzung von Social Media/WhatsApp 2020, ARD/ZDF-Onlinestudie vom 3. August 2021, online unter: <https://www.ard-zdf-onlinestudie.de/social-mediawhatsapp/> (abgerufen am: 14.07.2023).

2 Z. B. 24-Stunden-Inhalte, bei Twitter sog. „Fleets“, bei WhatsApp „Status-Meldung“, bei Instagram „Reels“ etc.

3 Sächsische Staatsregierung, Ministerien der Sächsischen Staatsregierung (Hrsg.), Coronavirus in Sachsen, 2021, online unter: www.coronavirus.sachsen.de (abgerufen am: 14.07.2023).

4 Library of Congress (Hrsg.), Update on the Twitter Archive at the Library of Congress, 2017, online unter: <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/> (abgerufen am: 14.07.2023).

5 Eine Umfrage zur praktischen Twitterarchivierung Anfang 2019 an eine Auswahl von Archiven im deutschsprachigen Raum, die sich mit der Archivierung von Netzressourcen beschäftigen, erbrachte eine Fehlanzeige.

6 Thomas Wildi, Archivieren aus Cloud-basierten Systemen. Die Beispiele Google Docs und Twitter, 2019, online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/bewertung-und-%C3%BCbernahme/08_Wildi_Praes_AUDS-Wildi_20190312_Wi.pdf (abgerufen am: 14.07.2023).

7 Peter Worm, Twitter Accounts von kommunalen Wahlbeamten. Bewertung und Überlieferungsstrategie, 2020, online unter: <https://archive20.hypotheses.org/8710> (abgerufen am: 14.07.2023).

8 Inkl. praktischer Übernahme: Florian Stabel, Sebastian Tripp, Hauptstaatsarchiv Wiesbaden übernimmt Twitter-Daten von Thorsten Schäfer-Gümbel, 2020, online unter: https://landesarchiv.hessen.de/twitter-daten_schaefer-guembel (abgerufen am: 14.07.2023).

9 Web Science and Digital Libraries Research Group auf Twitter unter <https://twitter.com/WebSciDL> (abgerufen am: 14.07.2023).

10 Z. B. Brendan Brown (Hrsg.), The Trump Twitter Archive V2, online unter: <https://www.thetrumparchive.com> (abgerufen am: 14.07.2023); The National Archives (Hrsg.), Donald J. Trump, Presidential Library, online unter: <https://www.trumplibrary.gov> (abgerufen am: 14.07.2023).

Pronom-Datenbank unter fmt/1311¹¹ (Droid-Signature-File V96) zwar gelistet, in der praktischen Formaterkennung aber instabil.

4 Fragen rund um die Twitterarchivierung

Die Bandbreite und Anzahl von sozialen Medien bzw. Netzwerken sind enorm.¹² Ressourcen anbietungspflichtiger Stellen erlauben mitunter nur eine kleine Auswahl dieser Kanäle. Bestenfalls liegt eine Social-Media- oder Kommunikationsstrategie zugrunde, was eine archivfachliche Bewertung erleichtert.¹³ Archivtechnisch können soziale Medien nicht pauschal bewertet werden, da Kanäle unterschiedliche Funktionen umfassen, unterschiedliche Daten produzieren bzw. zu archivierende Objekte zunächst zu definieren sind.¹⁴ Vergleichbar ist dies in Teilen mit der Archivierung von Inhalten aus Fachverfahren. Sind systemeigene Exportschnittstellen oder externe Tools bereits verfügbar, bietet sich ein Ansatz zur Übernahme. Als weiterer Aspekt ist die juristische Seite zu betrachten. Mögliche Leitfragen wären wie folgt: Können Inhalte Dritter, z. B. Kommentare, ohne Weiteres übernommen werden? Welchen Spielraum lassen AGB bei einer Datenerfassung (z. B. Crawlen vs. Scrapen)?

5 Twitter im Liveweb

Twitter („Gezwitscher“) ist ein Microblogging-Dienst (Start 2006), mittels dessen telegrammartige Kurznachrichten in Echtzeit veröffentlicht werden können. Die Posts sind dabei blogtypisch chronologisch in einer Timeline angeordnet. Anfangs umfasste ein Tweet max. 140 Zeichen, seit 2016 sind max. 280 Zeichen erlaubt. Charakteristika von Twitter sind Nutzer-Accounts (@Accountname), Follower-, Like-, Teil-Funktionen (Retweet), Kommentarmöglichkeit, Einbetten von Medien (Bilder, Videos, GIFs) und Links sowie Hashtags. Twitter dient als schnelles Nachrichtenmedium, aber auch, um Meinungen abzufragen oder Inhalte aus anderen Kanälen zu multiplizieren. Ein zielgruppenorientiertes Ausspielen von Content ist möglich. Dadurch, dass Inhalte beliebig geteilt werden können, ist nicht vorhersehbar, wer diese Inhalte insgesamt angezeigt bekommt.

6 Wanted: Ein Objekt zur Archivierung

Da Inhalte aus Systemen, aber nicht Systeme selbst archiviert werden, sind auf Basis signifikanter Eigenschaften zu archivierende Objekte zu definieren. Kategorien bilden z. B. „Aussehen“, „Funktionen“ sowie „Inhaltsumfang“. Twitter hat unterschiedliche Erscheinungsbilder je nachdem, ob es über einen Account angemeldet oder über das öffentlich einsehbare Liveweb (Public Time Line) angesehen wird. Ebenso spielt das responsive Layout eine Rolle (Desktop vs. mobil). Sollen (System-)Funktionen im Archiv nachgebildet werden? Zu kommentieren, teilen oder „ liken“ erübrigt sich, allerdings sind Recherche-, Filter- und Sortierfunktionen wünschenswert. Welche Inhalte wären zu übernehmen? Posts in ihrer Vollständigkeit¹⁵, Retweets, Kommentare, Metadaten, Medien (Bilder, Videos, GIF)? Schließlich ist zu entscheiden, ob ein Objekt als Timeline oder in Form einzelner Tweets oder Threads sowie ob es Account- oder Hashtag-bezogen zu archivieren wäre.

11 The National Archives (Hrsg.), Tweet JSON Signature, The Technical Registry Pronom, online unter: <https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=2129> (abgerufen am: 14.07.2023).

12 Hier geht es vorrangig um archivtechnische Belange.

13 Neben dem Hinzuziehen einer eventuell vorhandenen Social-Media-Strategie empfiehlt es sich darüber hinaus, mit dem jeweiligen Sozialen Medium an sich vertraut zu sein. Als Einstiegslektüre eignen sich z. B. Nicole Simon, Nikolaus Bernhardt, Twitter. Mit 140 Zeichen zum Web 2.0, München 2008; Tim O'Reilly, Sarah Milstein, The Twitter Book, Beijing et. al 2009; Anne Grabs, Karim-Parick Bannour, Eilsabeth Vogel, Follow me! Erfolgreiches Social Media Marketing mit Facebook, Instagram, Pinterest und Co., Bonn 2019, S. 381–409; Vivian Pein, Der Social Media Manager. Handbuch für Ausbildung und Beruf, Bonn 2019.

14 Hier sei nochmals auf Wildi hingewiesen, der eindrucklich über Definition von Objekten in cloudbasierten Anwendungen resp. Sozialen Medien und Schnittstellen referierte.

15 Im Liveweb kann ein Account-Inhaber jederzeit seine abgesetzten Posts löschen.

7 Erfassungsmethoden

Ein zur Erfassung von Websites gängiges Tool ist der Webcrawler, der als Objekt eine kopierte Website in zumeist statischer Form ausgibt. Für die Archivierung von Inhalten aus Twitter eignet sich diese Methode nicht, da Twitter zu dynamisch ist. Vereinfacht gesagt, erfassen Crawler nur die Inhalte, die am Bildschirm sichtbar sind. Bei Twitter sind die ersten paar Tweets am Bildschirm sichtbar, die Masse aber nicht.

Webbrowser bieten Funktionen zum Anlegen von Dateien, die sich aus Web-Content generieren: Funktionen wie „Speichern unter“ zur Objektanlage als „Website“, „Webarchiv“ (html), mit Webrecordern als AV-Datei oder einfachen Screenshots als Bilddatei. Diese Verfahren sind nicht massentauglich, sondern erfordern manuelles Ausführen oder zumindest Skripts zum Auslösen von Batch-Verfahren. So existieren inzwischen zahlreiche externe Tools, die Inhalte aus Twitter extrahieren bzw. kopieren können. Näheres dazu ist überblicksartig bei Worm und v. a. bei der Web Science and Digital Libraries Research Group zu finden.

Twitter bietet Account-Inhabern die Möglichkeit, sich sein „Twitter-Archiv“ auf „Knopfdruck“ auszugeben. Diese Funktion steht unbegrenzt (kostenfrei, beliebig wiederholbar) zur Verfügung. Bei Twitter wird das „Archiv“ angefordert und es wird in der Regel innerhalb weniger Stunden bis Tage als .zip-Paket per Download-Link bereitgestellt. Allerdings haben sich Datenstruktur und -umfang des Exports in den letzten drei Jahren bereits zweimal geändert. Folglich ist stetes Schnittstellen-Monitoring notwendig. Twitter gibt nur eigene Inhalte aus, das heißt, es werden keine Inhalte Dritter, sprich Kommentare, mitgegeben. Retweets sind nur verkürzt anbei.¹⁶

8 Verfahrensweise im Sächsischen Staatsarchiv

Der Account-Inhaber fordert das „Twitter-Archiv“ an. Das Exportpaket wird dem Archiv angeboten, übernommen, gegebenenfalls aufbereitet und archiviert.

Inzwischen wurden zwei Echtdatenübernahmen von verschiedenen Stellen durchgeführt. Die entsprechenden AIP sind archiviert und das Archivgut ist für eine Benutzung aufbereitet.

9 „Twitter-Archiv“-Datenanalyse

Seit Anfang 2018 werden regelmäßig Testexporte ausgeführt und Daten analysiert. Der erste Export stammt aus dem April 2018 und beinhaltet Rohdaten in JSON (Tweets monatsweise ausgegeben), zusätzlich alle Tweets und Retweets mit Metadaten in CSV, das Account-Profilbild sowie das Hintergrundbanner. Tweetmedien sind nicht vorhanden. Es wird eine Nutzeroberfläche in Form einer einfach gehaltenen, webbrowsersfähigen Timeline mit Such- und Kalenderfunktion bereitgestellt. In einer Übersicht (Readme.txt) kann ein Nutzer mehr über Paketinhalt und -struktur erfahren.

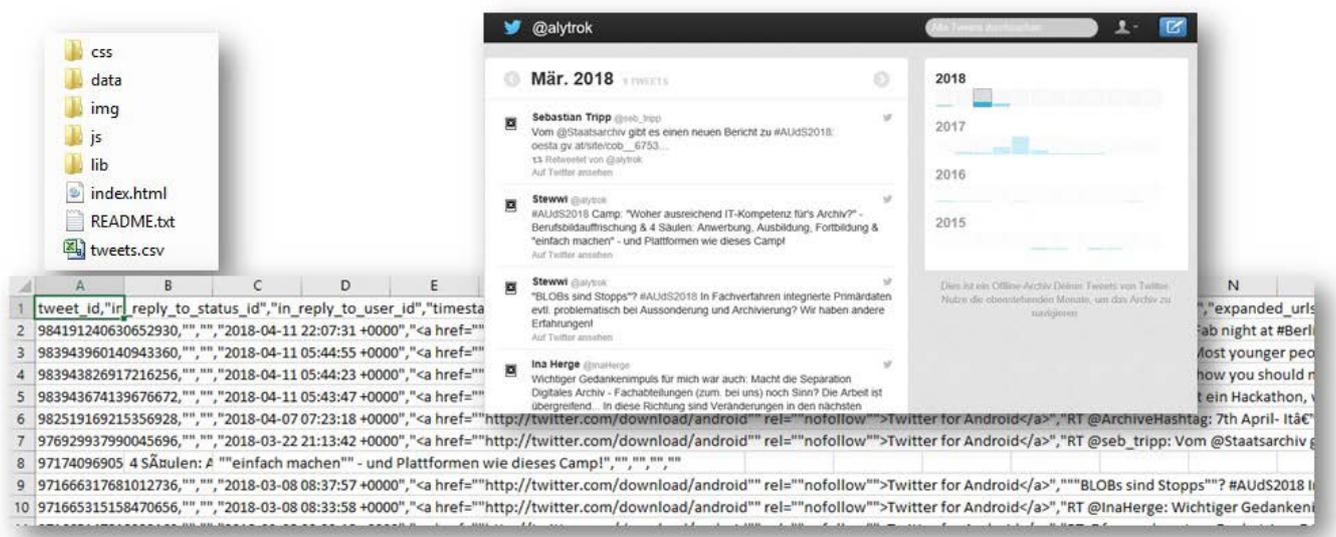


Abb. 1: Twitter-Archiv Stand April 2018. Eigener Screenshot.

¹⁶ Näheres zur Datenanalyse im späteren Verlauf.

Zwischen Juli 2019 und Januar 2020 beinhaltet ein „Twitter-Archiv“ wesentlich mehr Daten und ist detaillierter strukturiert. Sämtliche textbasierten Inhalte wie Tweets, Daten zum Account, Direktnachrichten, Likes, Retweets u. a. liegen in separaten JSON-Dateien vor. Medienobjekte (Bild, Video) aus eigenen Posts sowie Profilbild und -banner werden bereitgestellt. Bei AV-Medien können Objekte in verschiedenen Repräsentationen, genauer in verschiedenen Qualitätsstufen, vorliegen.¹⁷ Weiterhin liegt eine Übersicht (Readme.txt) vor.

Es wird kein browserfähiges GUI ausgegeben. Zum Lesen von Tweets kann zunächst nur die tweets.JSON verwendet werden. Allerdings sind Tweets darin nicht chronologisch angeordnet. Retweets sind nach einer bestimmten Zeichenkettenlänge abgeschnitten. Steuerzeichen sowie in Teilen Emojis sind in JSON umcodiert, sodass diese mit menschlichem Auge nicht unmittelbar erkennbar sind (Abb. 3).

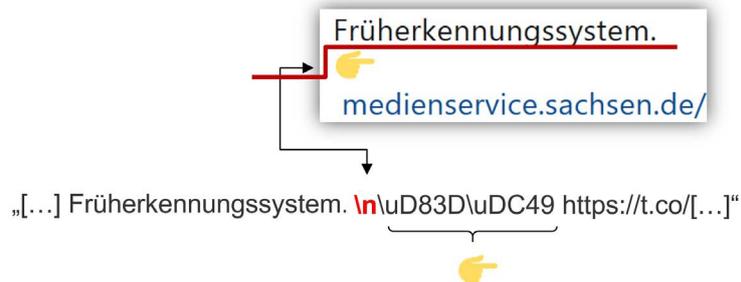


Abb. 2: Beispiel für Zeilenumbruch und Emojis in Liveweb-Twitter (oben) sowie in Twitter-Archiv in JSON Stand Juli 2019 (unten). Eigener Screenshot.

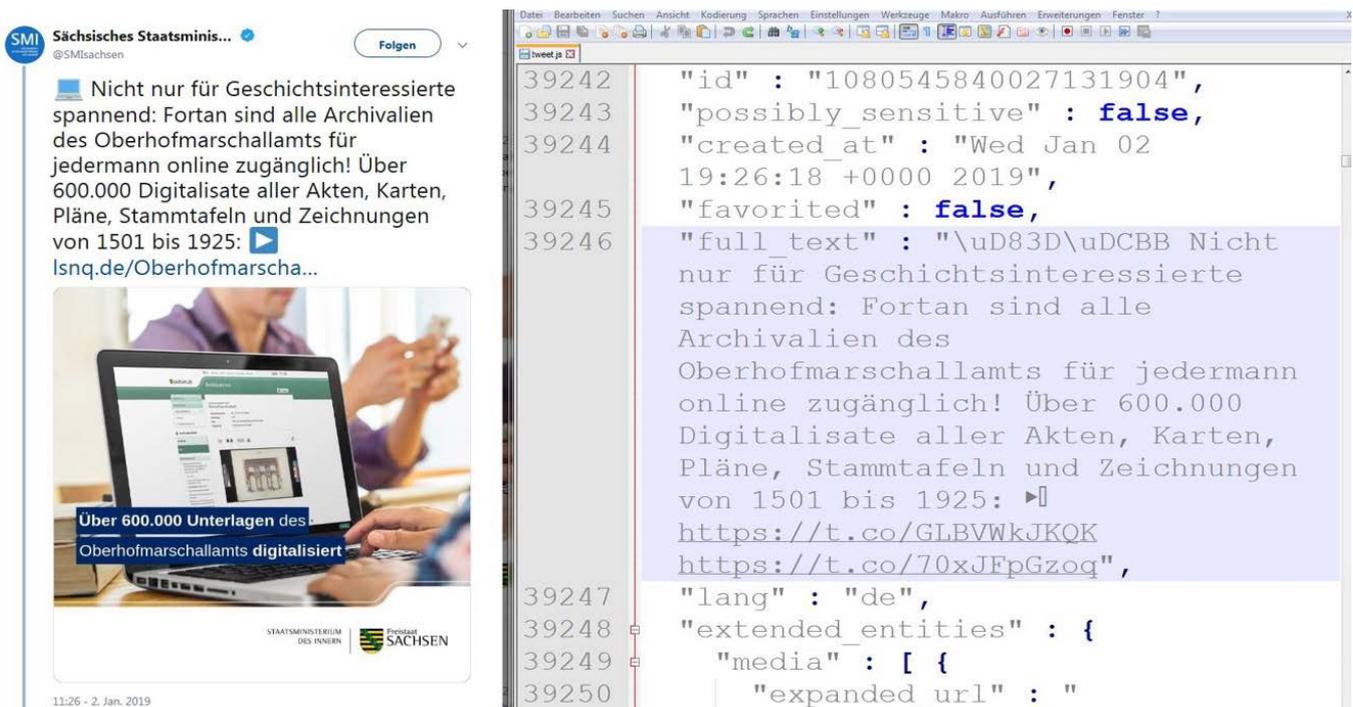


Abb. 3: Beispiel-Tweet im Liveweb-Twitter (links) und im Twitter-Archiv Stand Juli 2019 in JSON (rechts). Eigener Screenshot.

In den Zeitraum fällt die erste Echt Datenübernahme von Daten aus Twitter des Accounts des Sächsischen Staatsministeriums des Innern (SMI, @SMIsachsen). Das SMI twittert seit Juni 2014, hatte zum Zeitpunkt des Datenexports etwa 5.800 Tweets gepostet und über 4.000 Follower. Das „Twitter-Archiv“ umfasst etwa 1,8 GB gezippt, wovon die tweets.JSON lediglich 10MB umfasst. Während im Liveweb über die API nur die chronologisch letzten 3200 Einträge sichtbar sind, enthält das „Twitter-Archiv“ sämtliche Posts inklusive des ersten Tweets aus 2014.

17 So kann ein und dasselbe Kurzvideo in bis zu drei verschiedenen Qualitätsstufen im MP4-Format vorliegen. Möglicherweise liegt das an verschiedenen Endgeräten.

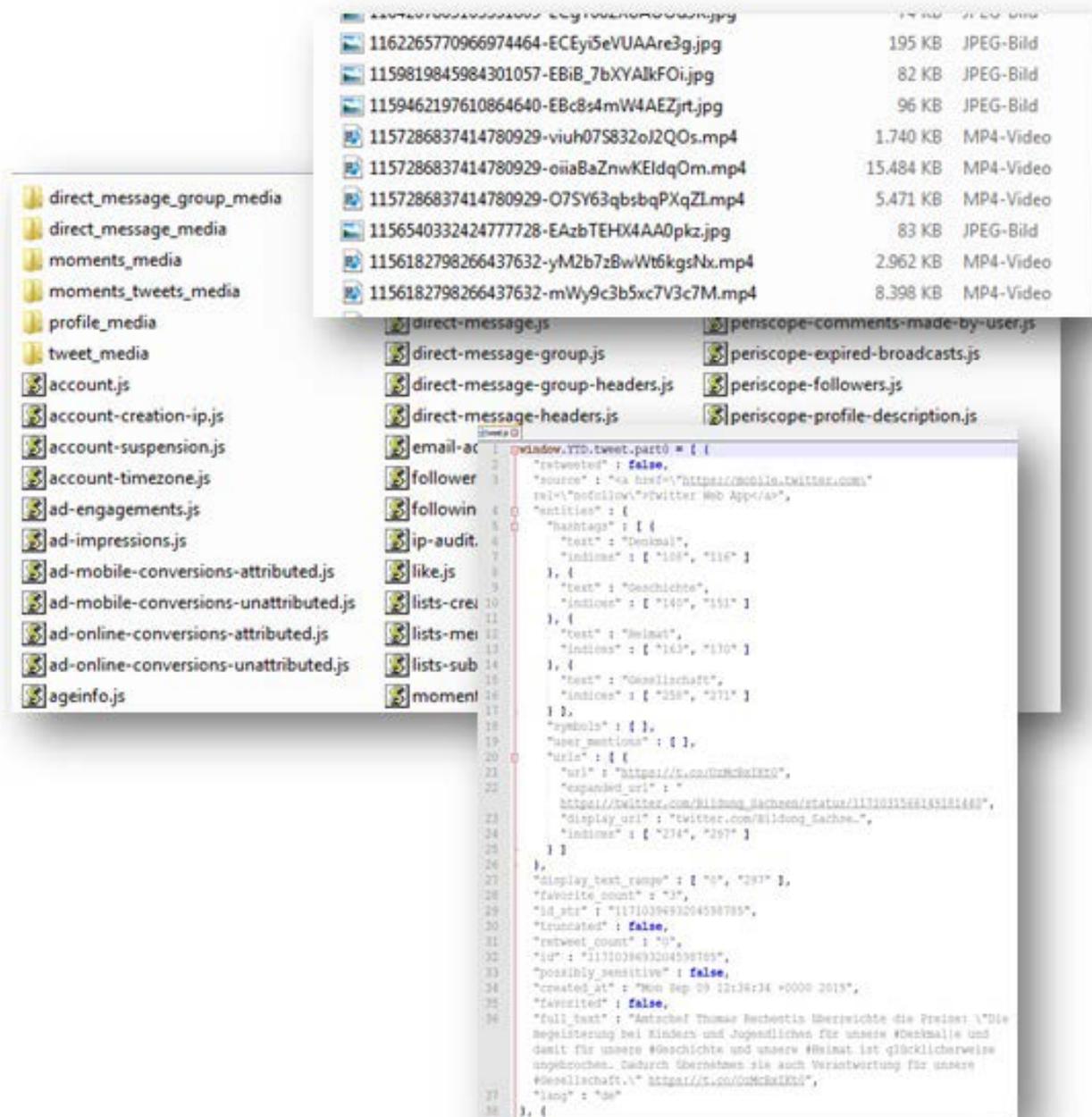


Abb. 4: Twitter-Archiv Stand September 2019. Eigener Screenshot.

Bei Exporten ab Januar 2020, welche u. a. einen weiteren Echtdatenexport über einen Account einer obersten Staatsbehörde umfassen, entspricht die Rohdatenstruktur den Exporten aus dem vorhergenannten Zeitraum. Hinzu kommt ein webbrowsersfähiges¹⁸ GUI, sodass ein Nutzer wie im Liveweb-Twitter mittels Timeline sowie mit Such- und Filterfunktionen im Datenbestand recherchieren kann.

Während Tweets weiterhin nicht chronologisch in der tweets.JSON vorliegen, Retweets weiterhin verkürzt ausgegeben werden, werden Emojis nun umfänglich visuell dargestellt.

In allen Exporten sind die ursprünglichen Links durch Twitter-eigene Kurzlinks überlagert. Im Quellcode der Webansicht bzw. in der JSON-Datei kann der jeweilige Ausgangslink gefunden werden. Außerdem fügt Twitter an vielen Stellen im Export Links auf Liveweb-Twitter zur statistischen Erfassung hinzu.¹⁹

18 Wie auch Livetwitter ist das „Twitter-Archiv“ nur mit bestimmten Webbrowsers kompatibel.

19 Der Twitter-eigene Kurzlinkdienst zeichnet sich durch „t.co“ aus.

10 Aufbereitung von Archivgut zur Nutzung

Während der letzte Echtdatenexport archivfähig ist und unverändert ingestiert werden konnte, mussten für den ersten Echtdatenexport Nutzungsvarianten geschaffen werden. So wurde zum einen das unveränderte „Twitter-Archiv“ übernommen. Zum anderen wurden dessen Rohdaten zusätzlich aufbereitet und dafür eine browserfähige Timeline rekonstruiert und ein Großteil der in JSON vorliegenden textbasierten Daten in eine Datenbank überführt (Abb. 6). Auf diesem Wege werden Look und Feel von Liveweb-Twitter weitestgehend nachempfunden. Die größte Herausforderung lag darin, die Tweets bzw. Retweets in eine chronologische Reihenfolge zu bringen, da die ursprüngliche Zeitstempelsyntax keine Sortierung zulässt, s. z. B. „Mon Sep 09 12:36:34 +0000 2019“ (Wochentag, Monat, Tag, Uhrzeit, Zeitzone, Jahr; alle Werte sind in einer Zeichenkette fortlaufend angegeben).

The image displays two main components related to the Twitter archive project:

- Twitter-Datenexport (Top):** A window showing the export options. It lists what is included (tweets, replies, retweets, likes) and what is excluded (replies to others, retweets of others). It also features a preview of a tweet from the account @SMISachsen, which includes a photo of a car crash and a logo for SMI (Sachsen Ministerium für Inneres).
- Database View (Bottom):** A screenshot of a database management tool (likely SQLiteSpy) showing a table named 'twitter_tweets'. The table has columns for 'id', 'created_at', 'full_text', 'truncated', 'favorited', 'favorite_count', 'retweeted', 'retweet_count', and 'lang'. The data is sorted chronologically. A pop-up window for 'SQLiteSpy - Edit Cell Data' is visible, showing a tweet about the digitization of historical records from the Oberhofmarschallamt.

Abb. 6: Aufbereitetes Twitter-Archiv als webbrowsersfähige Time Line (hinten) und in Form einer Datenbank (vorn). Eigener Screenshot.

11 Fazit zur Archivierung von Daten aus Twitter

Die Übernahmemethode über die Twitter-eigene Exportschnittstelle zur Erstellung eines „Twitter-Archivs“ bietet Vor- und Nachteile. Der Vorgang ist für Account-Inhaber kostenfrei, simpel und bis dato beliebig oft ausführbar. Tweets sind vollständig inklusive der Medienobjekte. Aktuell werden ein nutzerfreundliches GUI, umfangreiche Metadaten zum Account sowie Statistiken (z. B. Anzahl der Likes und Retweets) ausgegeben.

Dagegen wird in Kauf genommen, dass keine Kommentare anderer Nutzer ausgegeben werden. Retweets, sprich von Dritten übernommene Posts, werden verkürzt dargestellt.

Grundsätzlich muss aktiv Schnittstellen-Monitoring betrieben werden. Wie oben gesehen werden konnte, hat sich die Schnittstelle binnen weniger Jahre bereits zweimal verändert. Zwischenzeitlich mussten Exportrohdaten aufwendig aufbereitet werden, um eine komfortable Nutzung zu ermöglichen.

Im Gegensatz zu externen Tools zur Archivierung von Twitter spricht für die Variante über die Twitter-eigene Exportschnittstelle eine einfache Durchführung sowohl bei der abgebenden Stelle „auf Knopfdruck“ als auch beim Archiv aktuell ohne archivtechnische Nachbereitung. Auf diese Weise kann Vollständigkeit in Bezug auf die Anzahl der abgegebenen Tweets inklusive eingebetteter Medienobjekte erreicht werden. Werden hingegen externe Tools zur Erfassung

der Daten genutzt, die auf Live-Twitter im öffentlich einsehbaren Web (Public Time Line) ansetzen, werden hier nur jene chronologisch letzten 3.200 Einträge angezeigt, sodass mitunter ein Datenverlust einhergeht, wenn über einen Account bereits mehr als die besagte Anzahl an Posts veröffentlicht wurde. Außerdem ist der Aufwand zur Erfassung der Daten über separate Tools aufwendiger, da viele Schritte nicht im Batch-Verfahren ausgeführt werden können. Daten aus dem „Twitter-Archiv“ sind strukturell so weit aufbereitet, dass sie auch maschinell verarbeitet werden können (JSON). Tweets. JSON ist inzwischen in der Pronom-Datenbank aufgenommen und kann – wenn auch noch nicht stabil – durch Droid identifiziert werden.

Qualitätssicherung bei wenig verbreiteten Datenformaten am Beispiel von Digitalen Topographischen Karten

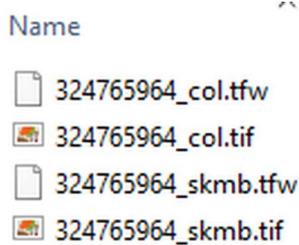
Oliver Brennecke, Antje Lengnik, Martin Vogel, Niedersächsisches Landesarchiv

1 Abstract

Aktuell existieren viele Möglichkeiten, Standard-Datenformate (PDF-, Bild-, Audio- und XML-Formate), die bei Aussonderungen aus Fachverfahren anfallen, mit einer Reihe von erprobten Softwarelösungen zu validieren (z. B. JHOVE, veraPDF, Fido, Jpylyzer). Es gibt hierzu viele Vorgaben und Vorschläge, ab wann Dateien eines Datenformats als valide und archivfähig angesehen werden. Doch wie validiert man wenig verbreitete Datenformate (wie sie z. B. bei Digitalen Topographischen Karten oder statistischen Mikrodaten verwendet werden), um die Qualität der Abgabe beurteilen und sichern zu können? Dieser Artikel beschreibt, wie solche Validierungen aussehen und die Ergebnisse in der Erschließung genutzt werden können.

2 Einleitung und Problembeschreibung

Im Niedersächsischen Landesarchiv (NLA) werden zurzeit digitale Geobasisdaten übernommen. Es handelt sich hierbei um Digitale Topographische Karten (DTK) in den Maßstäben 1:25.000, 1:50.000 und 1:100.000. Jeder Maßstab besteht aus einzelnen Kartenkacheln, die zusammen die gesamte Landesfläche abbilden. Die Abgabe umfasst folgende Struktur:



Name
 324765964_col.tfw
 324765964_col.tif
 324765964_skmb.tfw
 324765964_skmb.tif

Abb. 1: Verzeichnisstruktur der Abgabe

Zu jeder Kartenkachel existiert eine tfw-Datei sowie eine dazugehörige gleichnamige tif-Datei (z. B. 324765964_col.tfw und 324765964_col.tif). Die Kacheln liegen coloriert (_col) und in Graustufen (_skmb) vor, wobei nur die colorierte Version als archivwürdig übernommen wird. Die tfw-Datei¹ beschreibt die Georeferenzdaten der Kartenkachel mithilfe von UTM-Koordinaten. Zusätzlich enthält sie Informationen über die Breite und Höhe eines Pixels sowie die Rotations-terme für die x- und y-Achse. Die tif-Datei ist die grafische Repräsentation der Positionsinformationen.

Die folgende Abbildung zeigt eine grafische Repräsentation der Datei 324765964_col.tif. Die Originaldatei darf wegen noch laufender archivischer Schutzfristen nicht verwendet werden. Daher wird der äquivalente Ausschnitt aus OpenStreetMap² verwendet. Originalbeispiele sind auf der Webseite des Landesamts für Geoinformation und Landesvermessung Niedersachsen (LGLN)³ zu finden.

1 Wikipedia, World-File, online unter: <https://de.wikipedia.org/wiki/World-File> (abgerufen am: 14.07.2023)

2 OpenStreetMap, Kartenausschnitt, online unter: <https://openstreetmap.de/karte/?zoom=15&lat=53.84796&lon=8.65931&layers=B00TT> (abgerufen am: 14.07.2023).

3 Landesamt für Geoinformation und Landesvermessung Niedersachsen, Digitale Topographische Karten (DTK) und Übersichtskarten (DÜKN) als Rasterdaten – Beispiele, online unter: https://www.lgln.niedersachsen.de/startseite/geodaten_karten/topographische_geodaten_aus_atkis/rasterdaten_der_dtk_und_dukn/topographische-karten-uebersichtskarten-als-rasterdaten-beispiele-147360.html (abgerufen am: 14.07.2023).

die vorschlug, jede Kachel als separates AIP zu betrachten. Aus dem gesamten Zeitschnitt ein einziges AIP und damit auch einen einzigen Verzeichnungsdatensatz in Arcinsys zu generieren, bot sich allein schon aufgrund der schieren Menge an Kacheln – beim Maßstab 1:25.000 umfasst die Abgabe knapp 14.000 Stück – nicht an. Da ein Verzeichnungsdatensatz einer Bestelleinheit entspricht, würden Nutzende immer die gesamte Kachelanzahl erhalten, um sich dann mühselig die für sie relevanten herauszusuchen. Daher wurde entschieden, jede Kachel als separate Archivalieneinheit zu betrachten. Auch die vorliegenden Dateiformate waren bereits archivfähig und entsprachen den Empfehlungen der Leitlinie, sodass keine Konvertierung durch die Behörde durchgeführt werden musste.

Da die Daten abgesehen von der eindeutigen Nummer (UU_NO) der Kachel und den ausführlichen Koordinaten in der tfw-Datei keine beschreibenden Metadaten enthielten, musste eine Möglichkeit gefunden werden, die Kacheln aussagekräftig zu verzeichnen, um sie für Nutzende recherchierbar zu machen. Da es sich um sehr gleichförmige AIPs handelte, konnten schnell die Rumpfdaten der Erschließung festgelegt werden. Im Titel der Verzeichnungseinheit sollten der Produktname sowie die jeweilige Kachelnummer aufgenommen werden. Ergänzend wurden einige Textbausteine vorbereitet, die den Nutzenden technische Informationen über die enthaltenen Daten vermitteln, z. B. den inhaltlichen Aufbau einer tfw-Datei. Da im Archivinformationssystem Arcinsys gerade nach Karten aber oft anhand von Orts- und Gebietsbezeichnungen recherchiert wird, sollte jeder Verzeichnungsdatensatz im Idealfall zusätzlich die geografischen Angaben (Wohnorte, Gemeindegebiete, Gemarkungen) enthalten, die auf der jeweiligen Kachel abgebildet sind. Da es sich bei dieser Übernahme allerdings um mehrere tausend Kacheln handelt, kam eine händische Tiefenerschließung nicht infrage.

Glücklicherweise stellt die Landesvermessungsverwaltung (LGLN) auf ihrer Website ein sogenanntes Ortsverzeichnis bereit, das in Form einer csv-Datei alle Orts- und Wohnplatznamen Niedersachsens mit den dazugehörigen UTM-Koordinaten enthält.⁵ Außerdem konnten die Kachelraster der einzelnen Maßstäbe und die Verwaltungsgrenzen als Shape-Dateien zusätzlich übernommen werden.⁶ Mithilfe dieser zusätzlichen Informationsquellen boten sich Möglichkeiten zur automatisierten Anreicherung der Erschließungsdaten, welche in den weiteren Kapiteln näher beschrieben werden.

4 Validierung der DTK

Im folgenden Kapitel wird gezeigt, wie eine Validierung mit dem Dateiformat tfw durchgeführt werden kann. Hierzu ist es wichtig, die Begriffe Syntax und Semantik zu definieren.

4.1 Syntax:

Wenn wir über die Validierung von Syntax sprechen, ist in diesem Zusammenhang die Überprüfung von Eigenschaften auf Dateiebene gemeint.

Hierzu zählen u. a. die Formaterkennung (JHOVE, DROID, veraPDF), die Dateigröße (z. B. TreeSizeFree: Sind die Dateien leer?) usw.

4.2 Semantik:

Unter einer inhaltlichen Validierung (Semantik) verstehen wir die Analyse der Struktur der Abgabe, den Informationsgehalt der Daten (Dateiname, Inhalt der Daten usw.) sowie die Plausibilität der Daten (z. B.: sind die Koordinaten in der tfw-Datei UTM-Koordinaten?).

Das Problem bei der inhaltlichen Validierung besteht darin, dass es aktuell nur eingeschränkte bis gar keine Möglichkeiten gibt, Daten so ausführlich zu analysieren. Dies setzt zunächst einmal voraus, dass die Daten eine feste Struktur haben und Datenformate genutzt werden, die einem Standard entsprechen. Wenn keine solche Validierung erfolgt, kann es passieren, dass nicht nutzbare Daten archiviert werden (z. B. können Koordinaten in der tfw-Datei fehlerhaft sein). Beim tfw-Datenformat handelt es sich um einen international festgelegten Standard. Daher können die Daten nicht nur in Bezug auf ihre Syntax, sondern auch hinsichtlich ihrer Semantik validiert werden. Hierzu verfolgt das NLA einen eigenen Ansatz und hat einen Validator für dieses Datenformat entwickelt. Die aktuelle Version überprüft zwei syntaktische Eigenschaften der tfw-Dateien:

5 Landesamt für Geoinformation und Landesvermessung Niedersachsen, Ortsverzeichnis Niedersachsen (ATKIS), online unter: https://www.lgln.niedersachsen.de/startseite/geodaten_karten/topographische_geodaten_aus_atkis/ortsverzeichnis/ortsverzeichnis-atkis-144316.html (abgerufen am: 14.07.2023).

6 Landesamt für Geoinformation und Landesvermessung Niedersachsen, Verwaltungsgrenzen aus ATKIS, online unter: https://www.lgln.niedersachsen.de/startseite/geodaten_karten/topographische_geodaten_aus_atkis/verwaltungsgrenzen/verwaltungsgrenzen-aus-atkis-175378.html (abgerufen am: 14.07.2023).

1. Die Dateiendung lautet: *.tfw
2. Jede *.tfw Datei enthält genau (6) Zeilen, die nicht leer sein dürfen.

Da es sich um Positionsinformationen handelt, werden die UTM-Koordinaten (Zeilen 5 und 6) hinsichtlich ihres gültigen Wertebereichs validiert. Der Validator überprüft fünf semantische Eigenschaften des TFW-Datenformats:

1. Jede *.tfw Datei enthält eine Rechtskoordinate in Zeile (5), deren Wert zwischen 100.000 und 899.999 liegt.
2. Jede *.tfw Datei enthält eine Hochkoordinate in Zeile (6), deren Wert zwischen 0 und 10.000.000 liegt.
3. Es existiert zu jeder *.tfw Datei eine gleichnamige *.tif-Datei.
4. Es existiert immer eine farbige Kachel (*.tif_col) für eine *.tfw Datei“.
5. Die UTM-Koordinaten liegen in Niedersachsen.

Die folgende Abbildung zeigt die Benutzeroberfläche des Validators:

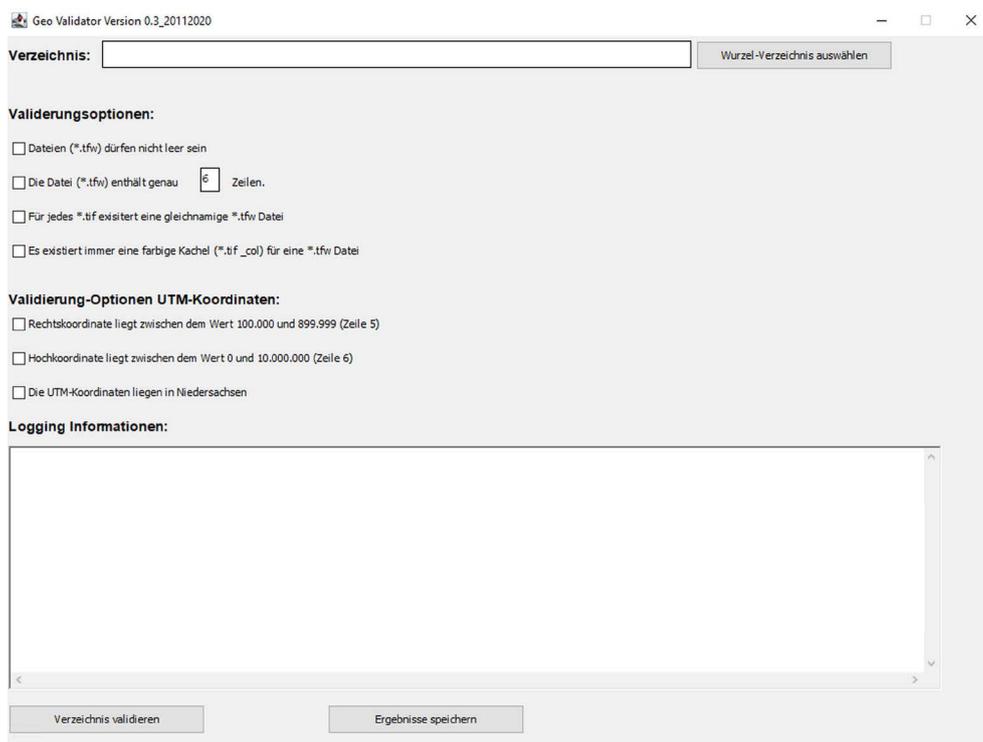


Abb. 3: Validator

Die Möglichkeit, zu überprüfen, ob die UTM-Koordinaten in Niedersachsen liegen, wurde genutzt, um weitere Meta-informationen (Wohnorte, Gemeinden, Gemarkungen) für die Erschließung zu erfassen.

5 Anreicherung der Verzeichnung mit weiteren Metadaten am Beispiel DTK 50

Die Grundlage für die Erschließung bilden die Daten, die als tfw-, tif- und Shape-Dateien vom LGLN geliefert werden. Die Shape-Datei beinhaltet die Koordinaten sowie die UU_NO (Abb. 8: Beispiel Cuxhaven) für die Kacheln. Weitere Informationen sind in der Shape-Datei nicht vorhanden.

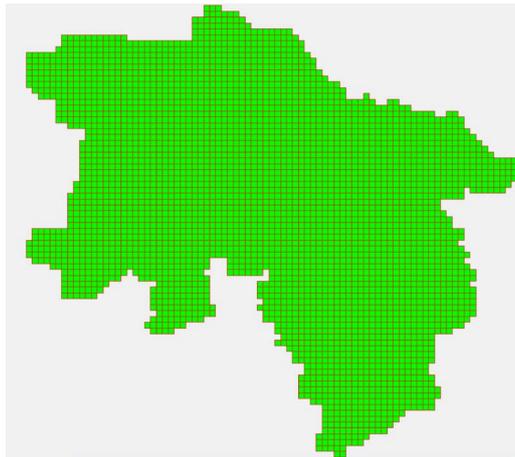


Abb. 4: Shape-Datei mit Kacheln

Die Metadaten der eigentlichen Kacheln enthalten, abgesehen von den Koordinaten, keine Informationen darüber, welche geografischen Gebiete jeweils auf ihnen abgebildet sind, wodurch sie im Archivinformationssystem nur bedingt recherchiert werden können. Es wurde daher beschlossen, die frei verfügbaren Informationen (Wohnorte, Gemeinden, Gemarkungen) des LGLN für die Anreicherung der Verzeichnungsdatensätze der Kacheln zu verwenden. Um diese Daten für die Verzeichnung nutzbar zu machen, wurden sie einheitlich in das Format GeoJSON⁷ konvertiert sowie eine eigene hausinterne Softwarelösung in JAVA geschaffen, um diese Metadaten verarbeiten und auf die Kartenkacheln mappen zu können.

Die nachfolgende Abbildung zeigt, welche Daten für das Mapping verwendet werden:

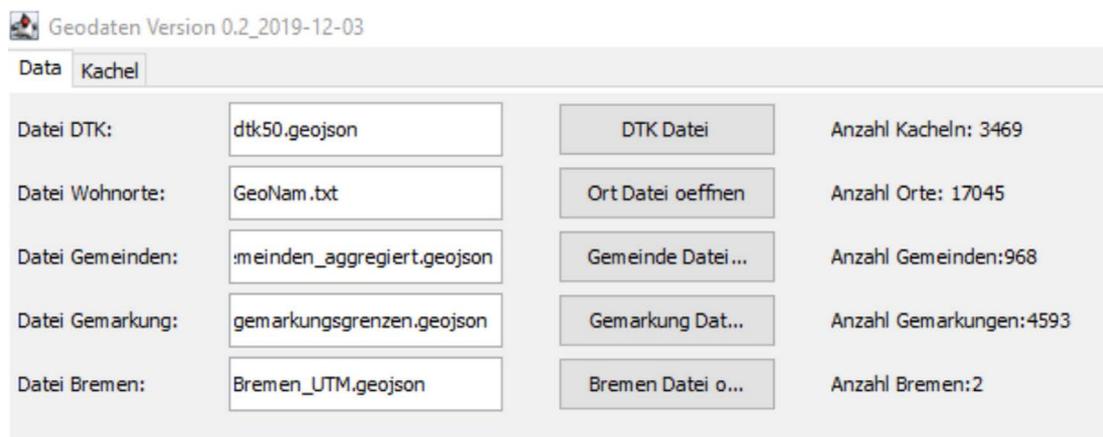


Abb. 5: Metadaten für das Mapping

Datei DTK: beinhaltet die Daten aus der Shape-Datei vom LGLN

Datei Wohnorte: Beinhaltet alle Wohnorte in Niedersachsen, die vom LGLN erfasst worden sind.

Datei Gemeinden: Beinhaltet alle Gemeinden, die in Niedersachsen erfasst sind.

Datei Gemarkung: Beinhaltet alle Gemarkungen, die in Niedersachsen erfasst sind.

Datei Bremen: Beinhaltet die Grenzen von Bremen. Dies ist wichtig, da die Kacheln auch das Landesgebiet von Bremen enthalten und dies bei der Erschließung mit angegeben werden soll.

Die nachfolgende Abbildung zeigt die verwendeten Daten (Gemeinden [Blau], Gemarkungen [Magenta], Bremen [Orange]) grafisch an. Das Kachelraster vom LGLN ist hier noch nicht zu sehen.

⁷ Wikipedia, GeoJSON, online unter: <https://de.wikipedia.org/wiki/GeoJSON> (abgerufen am: 14.07.2023).

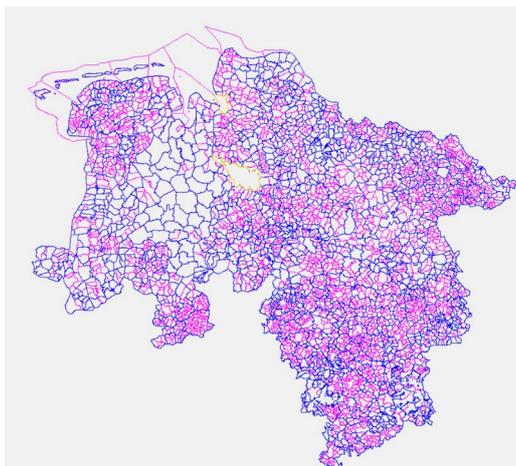


Abb. 6: Grafische Anzeige der Metadaten

Die nachfolgende Abbildung zeigt nun, welche Daten vom Kachelraster (ROT; Abgabe LGLN) erfasst werden:

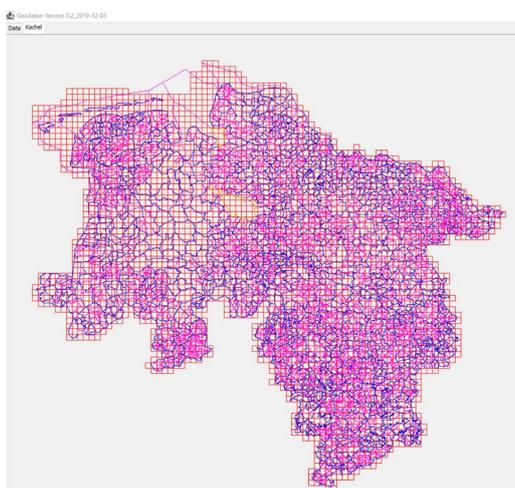


Abb. 7: Grafische Anzeige der Metadaten inkl. Kacheln

Wenn man nun in eine beliebige Kachel klickt, wird ermittelt, um welche Gemeinde es sich handelt und welche Kartenkachel (UU_NO) zugehörig ist. Im nachfolgenden Beispiel wurde Cuxhaven ausgewählt. Die blaue Fläche zeigt an, welche Kacheln zur Gemeinde Cuxhaven gehören. Die „Kachel Nr.“ gibt den Namen der zugehörigen tif-Datei aus. In diesem Fall ist es die Kachel mit der UU_NO: 324765964.

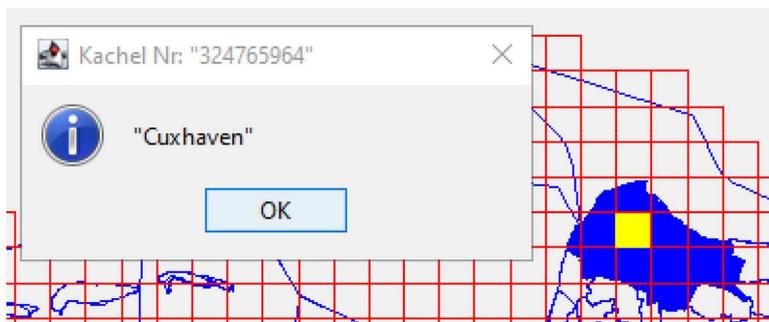


Abb. 8: Beispiel Cuxhaven

Das nachfolgende Bild zeigt einen vergrößerten Ausschnitt (Cuxhaven) mit den Gemeinden, Gemarkungen und Kacheln und des ausgewählten Bereichs an (grün) an.

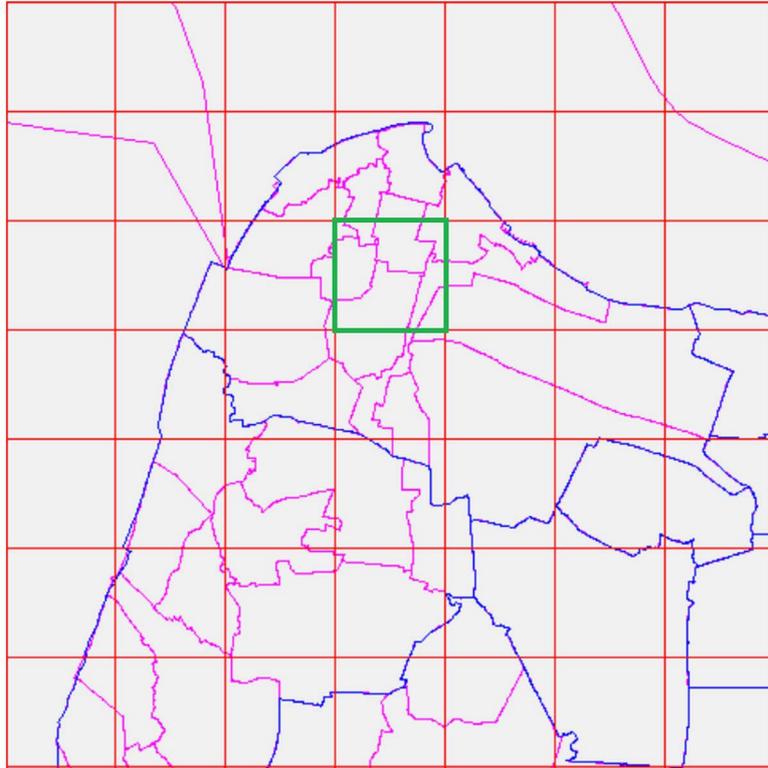


Abb. 9: Detaillierte Ansicht der Gemeinden und Gemarkungen

Der ausgewählte Bereich gehört zu der Kachel mit der Nummer 324765964 (siehe Abb. 2: Kartenkachel 324765964 in OpenStreetMap). Wie in Abb. 9: Detaillierte Ansicht der Gemeinden und Gemarkungen ersichtlich, werden hier mehrere Gemarkungen (Magenta, acht Stück) von der Kachel erfasst. Weiter kann man sehen, dass drei Orte in diesem Kachelausschnitt zu sehen sind (siehe Abb. 2: Kartenkachel 324765964 in OpenStreetMap). Das Mapping der einzelnen Metadaten (siehe Abb. 5: Metadaten für das Mapping) wird nun wie folgt durchgeführt:

1. Es wird für jede Kachel ein internes Shape-Objekt (grafisches Objekt in JAVA) erzeugt.
2. Es wird für jede Gemeinde (Blau) ein internes Shape-Objekt erzeugt.
3. Es wird für jede Gemarkung (Magenta) ein internes Shape-Objekt erzeugt.
4. Es wird für Bremen ein internes Shape-Objekt erzeugt.
5. Nun wird überprüft, ob das Shape-Objekt der Kachel aus 1. die Shape-Objekte der Gemeinden und Gemarkungen schneidet. Sollte dies der Fall sein, werden die Daten (Gemeinde, Gemarkung) erfasst.
6. Zuletzt wird überprüft, ob ein Wohnort (besteht nur aus einer X- und einer Y-Koordinate; also ein Punkt) im Shape-Objekt der Kachel liegt. Sollte dies der Fall sein, wird der Wohnort zu dieser Kachel erfasst.

Durch dieses Verfahren werden für jede Kachel die folgenden Metadaten erfasst, die automatisiert in den Enthält-Vermerk der Verzeichnungsdatensätze übernommen werden können:

UU_NO	Wohnplätze	Gemeindegebiete	Gemarkungen
324765964	Holte-Spangen, Stadt-/Gemeindeteil (Gemeinde: Cuxhaven, Stadt) Süder- und Westerwisch, Stadt-/Gemeindeteil (Gemeinde: Cuxhaven, Stadt) Süderwisch, Wohnplatz (Gemeinde: Cuxhaven, Stadt)	Cuxhaven	-Stickenbüttel (Gemeinde: Cuxhaven, Stadt) -Sahlenburg (Gemeinde: Cuxhaven, Stadt) -Holte-Spangen (Gemeinde: Cuxhaven, Stadt) -Altenwalde (Gemeinde: Cuxhaven, Stadt) -Franzenburg (Gemeinde: Cuxhaven, Stadt) -Süder- und Westerwisch (Gemeinde: Cuxhaven, Stadt) -Cuxhaven (Gemeinde: Cuxhaven, Stadt) -Groden (Gemeinde: Cuxhaven, Stadt) Altenbruch (Gemeinde: Cuxhaven, Stadt)

Durch diese zusätzlichen Metadaten können Nutzende gezielt in Arcinsys nach Wohnorten, Gemeinden und Gemarkungen suchen und bekommen die dazugehörigen Kacheln angezeigt.

6 Ergebnis

Dieser Artikel beschreibt die Möglichkeit, die Syntax und die Semantik von Geobasisdaten zu validieren und diese Ergebnisse für die Erschließung zu nutzen. Es wurden bei der Übernahme der Daten Kacheln identifiziert, die nicht valide hinsichtlich der Vorgaben aus Kapitel 3 sind. Die Probleme konnten jedoch behoben und dadurch die Datenqualität erheblich verbessert werden. Weiter können Nutzende durch die qualitativ hochwertige Erschließung in Arcinsys nun gezielt nach Orten, Gemarkungen und Gemeinden suchen. Ohne diese Metainformationen wäre eine Recherche nur anhand der Koordinaten möglich gewesen. Eine manuelle Erfassung dieser Metainformationen für jede einzelne Kachel wäre in Anbetracht der Größe der Abgabe nicht realisierbar.

Das NLA wird für weitere Abgaben des LGLN den Validator und das Mapping Tool verwenden und die Möglichkeiten des Mapping durch weitere Datenquellen erweitern.

Anonymisierung im Informationssystem des Historischen Archivs des ungarischen Staatssicherheitsdiensts. Umsetzung und Auswirkungen der COVID-19-Pandemie

Zoltán Lux, Historisches Archiv der Staatssicherheitsdienste, Budapest

Archive erweitern ständig das Leistungsspektrum in digitaler Form. Dies ist auch das strategische Ziel des Historischen Archivs der Staatssicherheitsdienste (Állambiztonsági Szolgálatok Történeli Levéltára, ÁBTTL). In diesem Archiv muss jedoch besonders darauf geachtet werden, dass sowohl beim zivilen Zugang als auch bei der wissenschaftlichen Forschung nur die ihnen zustehenden Informationen zugänglich sind und dass Forschende nur die relevanten Informationen in möglichst vollständiger Form erhalten. Der Entwicklungsplan für die kommenden Jahre zielte darauf ab, dies zu erreichen. Aufgrund der COVID-Pandemie musste der ursprüngliche Plan jedoch geändert werden. Einerseits musste schnell ein digitaler Online-Forschungsraum eingerichtet werden, andererseits musste - zusätzlich zum ursprünglichen Plan - eine Infrastruktur entwickelt und eingerichtet werden, in der die Mitarbeitenden des Archivs einen erheblichen Teil der Daten- und Dokumentenvorbereitung auch zu Hause im Homeoffice durchführen können.

1 Anonymisierung

Die beiden Haupttätigkeiten des Archivs sind die Bereitstellung der von den zeitgenössischen Geheimdiensten über die Bürger gesammelten Informationen an in der Zeit vor dem Regimewechsel beobachtete Personen auf der Grundlage der Informationskompensation und die Bereitstellung der im Archiv aufbewahrten Dokumente für die wissenschaftliche Forschung. In beiden Fällen müssen die Archive jedoch sicherstellen, dass die enthaltenen Sonderdaten (zum Beispiel Gesundheitszustand, religiöse Überzeugungen, Sexualverhalten, ethnische Herkunft usw.) ordnungsgemäß behandelt und nur den berechtigten Personen zugänglich gemacht werden. Daher müssen bei der Bereitstellung von Dokumenten in den Dokumenten bestimmte Teile abgedeckt und anonymisiert werden. Bei Papierdokumenten wurde dies durch doppeltes Fotokopieren und „Ablegen von Umschlägen“ ganzer Seiten im Archiv gelöst. Mit fortschreitender Digitalisierung der Akten wurde die Anonymisierung jedoch zunehmend digital. Im Jahr 2011 begannen wir mit verschiedenen Grafikbearbeitungsprogrammen zu experimentieren, wollten aber den Prozess „im System“ halten und andererseits eine Methode erarbeiten, bei der ein bestimmtes anonymisiertes Dokument „weiter modifiziert“ und verwendet werden kann, um zusätzliche anonymisierte Versionen zu erstellen.

Das Wesen der implementierten Anonymisierungsmethode besteht darin, auf dem von den Seiten zu bedeckenden Teil Rechtecke zu platzieren und die oberen linken und unteren rechten Koordinaten dieser Rechtecke für jede Seite in dieser Datenbank zu speichern (relativ zur oberen linken Koordinate der Bild).

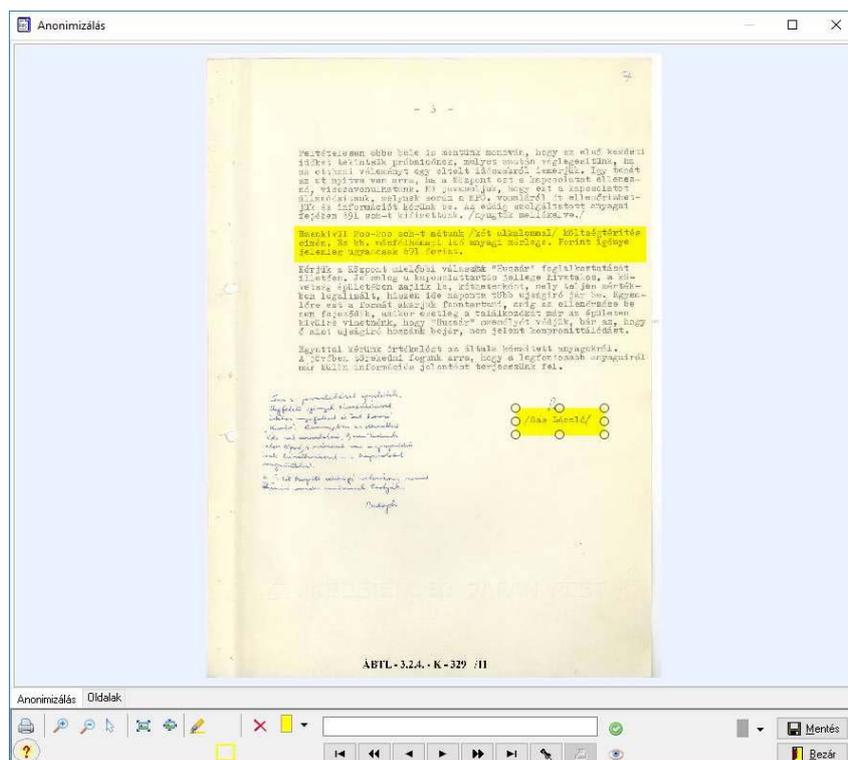


Abb. 1.: Anonymisierung einer Seite im internen Informationssystem.

Bei dieser Methode wird der Satz der oberen linken und unteren rechten Koordinaten an jede Bild-/Dokumentseite angehängt.

Ein Dokument (Datei, Dossier, einzelne Archivbeschreibungseinheit) besteht aus einer Reihe von (digitalisierten) Bildern.

Ein Dokument (eine Reihe von Bildern) hat eine (oder mehr) Anonymisierung(en), die aus einer Reihe von Koordinaten (oben links und unten rechts) besteht, die jeder Seite des Dokuments zugewiesen sind.

Natürlich gibt es Seiten von Dokumenten, die keine Informationen enthalten, die verschleiert werden sollen, weil sie entweder keine zu schützenden Daten enthalten oder in diesem speziellen Fall freigegeben werden können. Daher umfasst eine Anonymisierung auch eine „Gültigkeits“-Information für jede Seite des Dokuments, d. h. ob die Seite Teil eines bestimmten auszustellenden anonymisierten Dokuments ist.

Ein Dokument kann jedoch mehrere Anonymisierungen aufweisen, was bedeutet, dass ein Dokument mit einer Reihe von Rechtecken (Koordinaten) verknüpft ist, die die Seiten unterschiedlich bedecken.

Ein für Bürger*innen oder Forschende erstelltes anonymisiertes Dokument wird digital generiert, indem für jede Seite des Dokuments in der Datenbank geprüft wird, ob es zu dieser Anonymisierung gehört, und wenn ja, generiert die Software ein neues JPG-Bild basierend auf rechteckigen Koordinaten. In diesem neuen Bild werden die Pixel in den Rechtecken geändert (auf eine bestimmte Farbe). Diese Methode kann für Rasterbildformate verwendet werden. Aus den so anonymisierten Seiten (jpg-Bilder) erstellt die Software ein pdf-File, das keinen Zeichentext enthält.

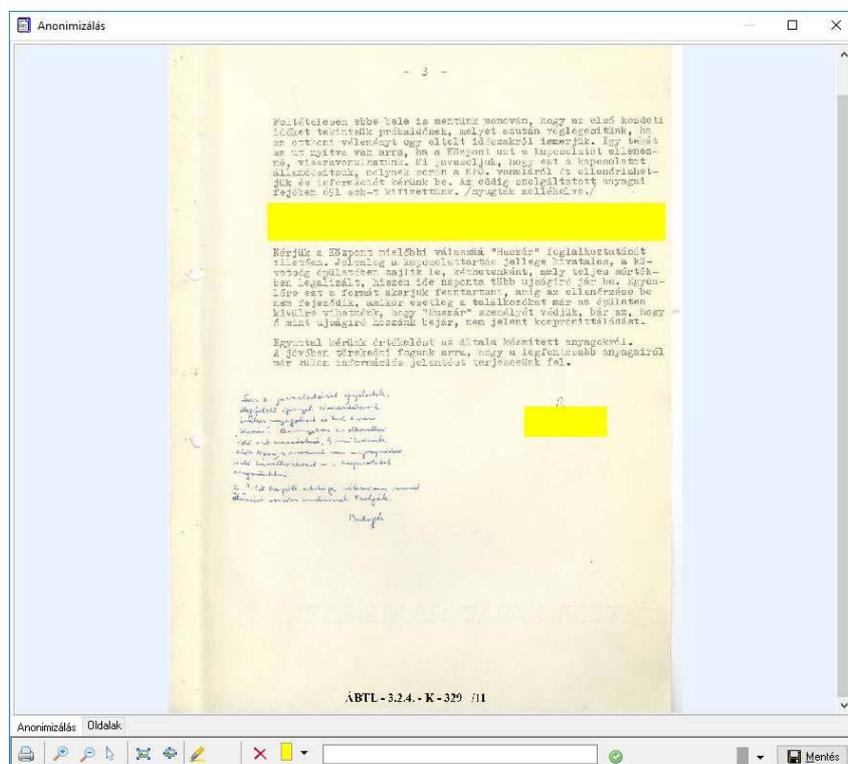


Abb. 2.: Eine anonymisierte Seite im internen Informationssystem.

Rechtecke eignen sich jedoch nicht nur zum Verdecken von Text in einem Dokument, sondern auch zum Hervorheben genau des Textes, den Sie dem Benutzer übergeben möchten. Dieser Fall ist von besonderem Interesse, wenn wir Informationen über Dokumente über eine bestimmte Person aus Dokumenten sammeln, die in den Archiven aufbewahrt werden, um Bürgereinsichten zu ermöglichen. In diesem Fall ist es nicht erforderlich, den Auskunftssuchenden die vollständigen Unterlagen zur Verfügung zu stellen, sondern nur die genauen Teile, die die Geheimdienste des sozialistischen Systems über die jeweilige Person gesammelt haben. Diese Textteile werden von den Archiven mit genauen Dokumenten- und Seitenkennzeichnungen gesammelt und den betroffenen Bürger*innen ausgehändigt. Rechtecke haben daher im Informationssystem ein zusätzliches Attribut, nämlich ob es als den Text „überdeckend“ oder als „zur Sammlung bezeichnend“ zu interpretieren ist. Natürlich können auch in dem mit „zur Sammlung bezeichnend“ gekennzeichneten Text Inhalte enthalten sein, die anonymisiert werden müssen. Daher haben Rechtecke noch eine weitere Eigenschaft: Ein Rechteck kann „immer verdeckt“ sein.

Ein weiteres Attribut von Rechtecken ist ihre Farbe. Dies scheint auf den ersten Blick keine wichtige Überlegung zu sein, aber wenn das Dokument entweder durch Forschende oder durch betroffene Bürger*innen auf Papier angefordert wird, ist es sehr wichtig, den Drucker „wirtschaftlich“ zu verwenden.

Bereits bei der Gestaltung der Anonymisierungsfunktion wurde vorgeschlagen, ein Rechteck, ein Polygon oder ein gedrehtes Rechteck zu verwenden, um einen Text zu verdecken oder auszuwählen. Dokumentbilder, die manchmal auf Originaldokumenten verzerrt sind, sind vertikale Texte. In vielen Fällen lässt sich ein Polygon durch den auszuwählenden Inhalt genauer abgrenzen als ein Rechteck. Aufgrund des Bedarfs an Entwicklungsressourcen und der praktischen Anwendbarkeit (es ist viel einfacher und schneller, ein Rechteck auf ein Bild zu „zeichnen“) wurde die „rechteckige Version“ jedoch mit der Absicht entwickelt, dass die Software weiterentwickelt wird, wenn tatsächlich Verbesserungsbedarf während der Verwendung entsteht. Nach den Erfahrungen der letzten zehn Jahre besteht jedoch nur in einem sehr geringen Teil der Fälle ein echter Bedarf an polygonaler Abdeckung.

Nachdem das Archiv anonymisierte Versionen der zu veröffentlichenden Dokumente (dies können anonymisierte Versionen vollständiger Dokumente oder Abbildungen ausgewählter Texte sein) für eine Recherche oder Einsicht erstellt hat, werden diese mit zweistufiger Genehmigung ausgestellt. Erst nachdem beide zuständigen Personen der Anonymisierung zugestimmt haben, können die anonymisierten pdf-Dateien einem Antrag auf Recherche oder Bürgerzugang zugeordnet werden, d. h. digital an den Auftraggeber ausgegeben oder analog ausgedruckt werden. Damit ist die Anonymisierung im Vorbereitungs- und Pflegeprozess vollständig in das Archivinformationssystem integriert, und das konkret ausgestellte Dokument wird auch innerhalb des Informationssystems archiviert und kann leicht für den Abruf oder andere damit verbundene Recherchen oder Erkenntnisse verwendet werden.

2 Homeoffice-Modul

Das interne Informationssystem des Archivs ist von außen nicht zugänglich. Als die COVID-19-Epidemie ausbrach, war es daher für die Archive ein großes Problem, ihren Mitarbeitenden Tätigkeiten im Homeoffice zu ermöglichen und die Nutzung durch wissenschaftlich Forschende nicht vollständig einzustellen.

Im Rahmen der Anonymisierung digitalisierter Dokumente für eine bestimmte Forschungs- oder Zugriffsanfrage werden Änderungen an einem genau definierten Teil der in der Datenbank gespeicherten Daten vorgenommen, mit denen andere Personen nicht gleichzeitig zu tun haben. Eine bestehende Anonymisierung wird in der Regel verwendet, um eine neue Anonymisierung eines Dokuments für andere Recherche- oder Betrachtungszwecke zu erstellen. Die Digitalisierung der Archive beträgt weniger als 10 % des Gesamtbestandes. Um den Nutzungsbedürfnissen gerecht zu werden, ist es notwendig, weit mehr vollständig neue anonymisierte Dokumente zu erstellen, als wir eine neue Anonymisierung basierend auf einer bestehenden erstellen können. Das Anlegen neuer anonymisierter Versionen jedes Dokuments ist größtenteils eine Aufgabe, bei der Datenänderungen im internen Informationssystem des Archivs je nach Nutzungsvorgang leicht unterschieden werden können. Diese Datenänderungen sind neue Einträge in Datentabellen, die erstellt wurden, um die Anonymisierungsfunktion zu unterstützen. Der Koordinatensatz der Rechtecke muss in eine Datentabelle eingefügt und dann digitalisierten Dokumentenseiten zugeordnet werden, die bereits in das Informationssystem geladen sind. Die eindeutigen Identifikationsdaten (IDs) dieser Dokumente und Dokumentenseiten können aus dem internen Informationssystem exportiert und in eine lokale Datenbank geladen werden. Dabei können Nutzer eine Anonymisierung wie in einem internen Informationssystem vornehmen. Rechtecke werden auf den zu überdeckenden / zu markierenden Text gelegt, sie erhalten die entsprechenden Attribute (immer verdeckt, Farbe). Auf den abzudeckenden oder zu markierenden Text werden Rechtecke mit den entsprechenden Attributen (immer verdeckt, Farbe) platziert. Die Struktur der Datentabelle, die die Koordinaten in der lokalen Datenbank aufzeichnet, entspricht daher genau der Struktur in der internen Datenbank.

Die Heimanonymisierung beginnt mit dem Export der wichtigsten Metadaten, digitalisierten Seiten (jpeg-Format) und Metadaten des zu anonymisierenden Dokuments aus der internen Datenbank, der von jedem autorisierten Benutzer durchgeführt werden kann. Diese Exportdateien können in das lokale Programm (Homeoffice-Modul) geladen werden, wo Benutzer neue Anonymisierungen erstellen können.

Die abgeschlossenen neuen Dokumentenanonymisierungen werden von Benutzern (aus dem Homeoffice-Modul) exportiert und dann in das interne Informationssystem geladen, wo die Ausgabe anonymisierte PDF-Dateien (die nur Bilder enthalten) generiert und der entsprechenden Recherche- oder Insight-Aktion zugeordnet werden. Im Homeoffice-Modul ist es auch möglich, anonymisierte PDFs zu generieren, da die mit der Übermittlung verbundene Verwaltung jedoch im internen Informationssystem erfolgen muss, erfolgt dies im internen System.

3 E-Forschung

Die ÁBTL hat seit einigen Jahren eingeführt, dass bereits digitalisierte Dokumente den Forschenden digital (und ordnungsgemäß anonymisiert) innerhalb des Archivs zur Verfügung gestellt werden (davon kann nur in sehr seltenen, eigens begründeten Fällen abgewichen werden). Dies sicher online zu tun, ist keine leichte Aufgabe.

Einerseits, weil sichergestellt werden muss, dass das digitalisierte Dokument nur autorisierten Personen zugänglich ist. Neben den schriftlichen Stellungnahmen der Forschenden überwacht das System permanent die Logins und Aktivitäten der Nutzer und sperrt den Nutzer bei einem „ungewöhnlichen“ Versuch (direkte Zugriffsversuche, mehrfache Login-Versuche von entfernten Standorten).

Zum anderen, weil das Dokument nur dann zum Download zur Verfügung stehen soll, wenn es bestellt wird. In diesem Fall ist es nicht die Gebühr für die Dienstleistung, die dem Archiv wichtig ist, sondern dass diese Dokumente die Institution nicht ungeprüft verlassen. Daher kann ein anonymisiertes Dokument, das einer Person im Internet nur zu Recherchezwecken zur Verfügung steht, nicht heruntergeladen werden und eine Seite ist nie vollständig auf dem Bildschirm zu sehen, sie kann nur „unter ein Fenster“ verschoben werden.

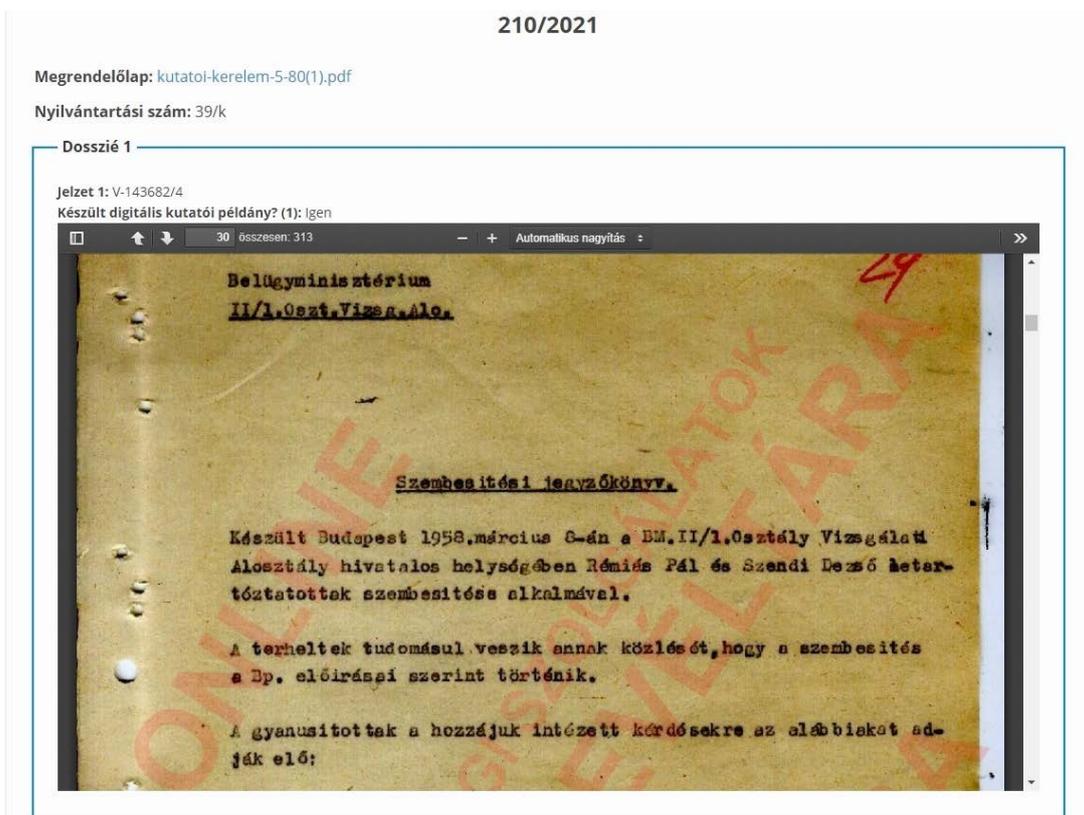


Abb. 3.: Anzeige des anonymisierten Dokuments auf der Online-Rechercheoberfläche.

Während die meisten Ereignisse dieser Art bisher darauf zurückzuführen waren, dass Personen Probleme hatten, ihre Passwörter zu ändern oder sie vergessen hatten, gab es bereits Nutzende, die direkte Downloads versuchten.

4 Zusammenfassung

Der Beitrag stellt die Methode der digitalen Anonymisierung vor, die im Historischen Archiv des Staatssicherheitsdienstes seit 10 Jahren zur Aufbereitung von Dokumenten für Recherchen und Bürgereinsichten verwendet wird, und deren Einbindung in das Archivinformationssystem. Die Anonymisierung digitalisierter Dokumente ist ein relativ eigenständiger Vorgang, andere Metadaten der Dokumente werden nicht verändert. Die Existenz dieser Funktion, während der COVID-19-Epidemie war in zweierlei Hinsicht sehr nützlich. Es war relativ einfach, eine eigenständige Anwendung zu entwickeln, die es Archivmitarbeitern ermöglicht, zu Hause, im Homeoffice zu anonymisieren und damit die für die E-Recherche benötigten Inhalte zu erstellen. Die Entwicklung eines Frameworks für e-Research war bereits eine große Aufgabe, die aber in den Entwicklungsplänen für die kommenden Jahre enthalten ist.

Vernetzen, Erschließen, Verwalten, Zugänglichmachen. Die Kulturerbedatenbank der Deutschsprachigen Gemeinschaft Belgiens.

Vitus Sproten, Belgisches Staatsarchiv, Eupen

1 Einleitung



Abb. 1: ZOG, Bestand Martin Schröder, Paris 1977 Ausflug des Kegelclubs (CC BY-NC-ND 4.0)

Ein Bild des Kegelclubs „*Rührt Euch*“ aus Eupen, das wahrscheinlich im Frühjahr 1977 aufgenommen wurde. Martin Schröder, Urheber des Bildes, aus dem kleinen Dorf Valender in der belgischen Eifel gehörte diesem Klub an. Sein Bruder Erich erinnert sich an die Aufnahme des Bildes: „Das Foto wurde mit ziemlicher Sicherheit auf der Aussichtsplattform des *Arc de Triomphe* gemacht. Der Kegelclub wurde 1969 von hauptsächlich in Eupen wohnenden Eifelern gegründet. Danach sind auch ein paar Eupener hinzugekommen. Bis 1986 haben wir zusammen gekegelt. Dann wurde aus dem Kegelclub ein Skatclub. Der wurde dann Ende 2010 aufgelöst.“ Das Bild zeigt neben Erich Schröder weitere Mitglieder des Kegelclubs, die alle aus dem Gebiet deutscher Sprache Belgiens stammten.

Neben diesen Informationen sind der Aufnahme in der sogenannten *Kulturerbedatenbank der Deutschsprachigen Gemeinschaft* (im Folgenden auch kurz „Kulturerbedatenbank“) eine Bestandsnummer, eine Rechtfreigabe, ein Bildurheber, Informationen über das Archiv und zahlreiche weitere Metadaten beigefügt.¹ Öffentlich abrufbar ist das Bild über das Geschichtsportale des Zentrums für Ostbelgische Geschichte, das an die *Kulturerbedatenbank* angeschlossen ist.

Das hier angeführte Beispielbild zeigt, dass sich in den 2010er- und 2020er-Jahren Archive nicht mehr nur noch Gedanken über das technische „Wie“ bei der digitalen Archivierung von Archivgut machen, sondern dass verstärkt auch das inhaltliche „Was“ und die Zugänglichkeit von digitalem Archivgut in den Interessenfokus von Verwaltungen, Archivaren, Historikern, Kulturschaffenden und Geschichtsinteressierten rückt.²

Diese Entwicklung lässt sich exemplarisch an den Veränderungen der *Kulturerbedatenbank* nachzeichnen, die sich von einer reinen Datenbank zur Verwaltung des Denkmalkatasters des Gebiets deutscher Sprache Belgiens hin zu einem flexiblen Tool zur Erschließung, Verwaltung, Zugänglichmachung und Zwischenarchivierung von Archivalien und Kulturgut entwickelt hat, wobei ein besonderer Fokus auf eine „bottom-up“ Bestandsbildung und Erschließung gelegt wurde. Während die Plattform in einem ersten Schritt im Auftrag der Deutschsprachigen Gemeinschaft Belgiens entwickelt, finanziert und betreut wurde, sind es nunmehr zahlreiche Partner und Projektteilnehmer, die die Datenbank

1 Pelle Snickars, *If Content is King, Context is its Crown*, in: *VIEW Journal of European Television History and Culture* 1 (2012), S. 34–39.

2 Vgl. hierzu auch die Beiträge von Edouard Vasseur und Annette Birkenholz im vorliegenden Band. Exemplarisch lassen sich ebenfalls die zahlreichen Projekte im Bereich des *Crowdsourcings* anführen, die ebenfalls zu dieser Interessenverlagerung beitragen. Vgl. hierzu auch Kate Theimer, *Partizipation als Zukunft der Archive*, in: *Der Archivar* 1 (2018), S. 6–12; hierzu trägt zweifelsohne die Perspektive der Digital Humanities auf diese Fragen bei: Andreas Fickers, *Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik*, in: *Zeithistorische Forschung* 12 (2020), S. 157–168; Andreas Fickers, Valérie Schafer, Christine Barats, *Fading Away... The challenge of sustainability in digital studies*, in: *Digital Humanities Quarterly* 14 (2020), online unter: <http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html> (abgerufen am: 14.07.2023)

nutzen und Archivalien in das System einpflegen. Dabei legt das Projekt in der kritischen Gesamtschau aber zweifelsohne auch Schwachstellen offen, die bei einer zu kurzfristigen Planung und den hiermit verbundenen Pfadabhängigkeiten entstehen können. Im Folgenden soll daher auf die Entwicklungsstufen der *Kulturerbedatenbank* eingegangen werden, welche Möglichkeiten diese für die digitale Archivierung bietet und welche Potenziale sie in der Zukunft entwickeln könnte.

2 Institutionelle Rahmenbedingungen

Das Gebiet deutscher Sprache Belgiens – offiziell Deutschsprachige Gemeinschaft Belgiens (im Folgenden auch „Ostbelgien“) – verfügt seit 1973 über eine kulturelle Autonomie im belgischen Staatsgefüge, die sich bis heute im Rahmen von sechs Staatsreformen allmählich ausgeweitet hat und neben kulturellen, personenbezogenen Materien – etwa das Gesundheitswesen oder die Kulturpolitik – nunmehr auch regionale Befugnisse – etwa Raumordnung oder Beschäftigungspolitik – umfasst.³

Das Gebiet verfügt dementsprechend über eine eigene Legislative und seit 1983 ebenfalls über eine Regierung und Verwaltung. Hierbei ist der „Fachbereich Kultur“ des Ministeriums der Deutschsprachigen Gemeinschaft für den Bereich Kultur zuständig. Die Kompetenzen dieses Fachbereichs erstrecken sich von der Archäologie, über die Bezuschussung von Veröffentlichungen, die Gestaltung des Sammlungsmanagements der Kunstsammlung der Deutschsprachigen Gemeinschaft bis hin zum Denkmalschutz. Eben jener Fachbereich entwickelt seit den 2000er-Jahren auf Verwaltungsseite auch die Leitlinien der hier behandelten *Kulturerbedatenbank*.

Hierbei zeugt die Integration des Projektes in das Regionale Entwicklungskonzept (REK) Ostbelgiens von der politischen Unterstützung, auf die sich die *Kulturerbedatenbank* stützen kann. Im Regionalen Entwicklungskonzept entwarf die Regierung der Deutschsprachigen Gemeinschaft bis dato drei (2009–2013, 2014–2019, 2019–2024) mittelfristige Zukunftsvisionen für unterschiedliche Projekte, die das Leben der Bürger*innen in Ostbelgien nachhaltig beeinflussen sollen.⁴ Als ein wichtiges Handlungsfeld wurde hierbei die Vermittlung des kulturellen Erbes der Region ausgemacht, wobei die *Kulturerbedatenbank* als wegweisendes Instrument zur Neu-Inwertsetzung und digitalen Sicherung dieses Kulturerbes dienen soll.⁵

Sowohl auf Verwaltungsebene als auch aufseiten der Regierung bestand dementsprechend der Wille, eine digitale Plattform zur Speicherung des kulturellen Erbes aufzubauen.

3 Entwicklungsstufe 1 (2005–2011): Vom Denkmalkataster zur Datenbank

Bevor das Projekt der *Kulturerbedatenbank* angegangen wurde, ging von der Denkmalschutzabteilung des Ministeriums der Deutschsprachigen Gemeinschaft in den beginnenden 2000er-Jahren die Initiative aus, ein neues, umfangreiches Denkmalkataster der Deutschsprachigen Gemeinschaft anzulegen und dieses auch digital zur Verfügung zu stellen.⁶ Obwohl ab 2003 bereits die Nutzung der „neuen Medien“ zur attraktiven Darstellung des Katasters ins Auge gefasst wurde, erwies sich das im Jahr 2009 vorgestellte Endprodukt noch als wenig benutzerfreundlich. Als Manko erwies sich insbesondere die Tatsache, dass die Informationen des digitalen Buches „[...] ähnlich wie auf einer langen Schriftrolle hintereinander aufgelistet waren“.⁷

3 Vgl. zur Geschichte des Gebiets deutscher Sprache in Belgien unter anderem Christoph Brüll, Carlo Lejeune, Grenzerfahrungen, Eine Geschichte der Deutschsprachigen Gemeinschaft Belgiens. Band 5: Säuberung, Wiederaufbau, Autonomiediskussionen (1945–1973), Eupen 2014; Katrin Stangherlin (Hrsg.), La Communauté germanophone de Belgique – Die Deutschsprachige Gemeinschaft Belgiens, Brügge 2005.

4 Vgl. hierzu Ministerium der Deutschsprachigen Gemeinschaft Belgiens (Hrsg.), DG – Ostbelgien Leben 2025. Regionales Entwicklungskonzept der Deutschsprachigen Gemeinschaft, Band 1–5, Eupen 2011–2019.

5 „Die Digitalisierung des Kulturerbes ist ein Schlüsselbereich der am 9. April 2019 durch die EU-Staaten unterzeichneten Erklärung zur Gewährleistung der digitalen Zukunft. Das reiche kulturelle Erbe Europas verkörpert unsere Vergangenheit und inspiriert uns mit dem Gefühl einer gemeinsamen Geschichte und Identität. Durch die Digitalisierung des Kulturerbes in Ostbelgien mittels der Datenbank „Kulturerbearchiv“ werden Bestände der Museen, Geschichtsvereine, Archive und der öffentlichen Hand systematisch digital erfasst. Ziel ist die langfristige Sicherung der wichtigen Zeugnisse unserer Vergangenheit und ihre digitale Zugänglichkeit.“, online unter: https://ostbelgienlive.be/PortalData/2/Resources/downloads/rek/Regionales_Entwicklungskonzept_Band_5_REK_III.pdf (abgerufen am: 14.07.2023)

6 Die Kompetenz Denkmalschutz wurde zunächst (1983–1988) durch die Deutschsprachige Gemeinschaft ausgeführt, hiernach (1988–1994) wurde diese Kompetenz an die Wallonische Region übertragen, um dann seit 1994 wieder durch die Deutschsprachige Gemeinschaft ausgeführt zu werden. Vgl. für die rechtliche Grundlage des Denkmal- und Landschaftschutzes auch das Denkmalschutzdekret vom 23. Juni 2008.

7 Grenz-Echo, Die Tat ist vergangen, aber die Denkmäler bleiben, 17. März 2009; vgl. zu den Vorplanungen ebenfalls Grenz-Echo, Kataster der schützenswerten Gebäude und Landschaften, 19. Dezember 2003.

Auf diesem ersten digitalen Denkmalkataster aufbauend, entschied das Ministerium der Deutschsprachigen Gemeinschaft im Jahr 2008, eine Datenbank zur digitalen Erfassung, Erschließung und Nutzbarmachung für ihr umfangreiches Kulturerbe einzurichten⁸. Diese sollte es außerdem in Zukunft erlauben, die Informationen aus der Datenbank weniger statisch darzustellen. Anberaumt wurde, dass diese Datenbank nicht mehr nur noch Informationen zum Denkmal- und Landschaftsschutz, sondern ebenfalls Angaben und digitalisierte Quellen zur Kunstsammlung der Deutschsprachigen Gemeinschaft oder zu archäologischen Stätten beinhalten sollte. Desgleichen wurde für die Datenbank die Erfassung des immateriellen Kulturerbes – etwa Bräuche oder Dialekte – vorgesehen. In der Folge entstand eine Datenbank, die auf die Bedürfnisse der Kulturabteilung des Ministeriums maßgeschneidert war und insbesondere den Wünschen der Abteilung Denkmalschutz und Archäologie Rechnung trug. Ein Hauptaugenmerk lag deshalb bei der ersten Entwicklungsphase unter anderem auf der elektronischen Verwaltung der Akten der Denkmalschutzabteilung des Ministeriums.

Hierbei wurde als Grundlage für die Datenbank ein Objektframework genutzt, das auf Microsoft.NET basiert und von einem regionalen Anbieter unter dem Titel „Demotec Objects“ entwickelt wurde. Diese relationale Objektdatenbank dient bis heute als Grundlage für die *Kulturerbedatenbank* der Deutschsprachigen Gemeinschaft.⁹

Dass die ersten Adressaten der Datenbank am Ende der 2000er-Jahre insbesondere Kunsthistoriker, Archäologen und Verwaltungsangestellte waren, lässt sich bis heute in der Entwicklung und den Schwerpunkten des Projekts ablesen. In der Datenbank standen bzw. stehen nämlich einzelne Objekte (Datenbankeinträge) zentral und alle weiteren, dazugehörigen Elemente (etwa digitalisierte Quellen) wurden ihnen untergeordnet. Hiermit wurde insbesondere die Wissensstrukturierung durch das Fachpersonal der Verwaltung berücksichtigt.

Konkret wurde in dieser ersten Projektphase zu jedem Denkmal oder Kunstwerk im Besitz der Deutschsprachigen Gemeinschaft und jeder archäologischen Grabungsstätte im Gebiet ein Datenbankeintrag erstellt. Das heißt pro denkmalgeschütztem Objekt existierte von nun an ein Eintrag in der Datenbank.¹⁰ Diese Einträge wurden durch Beschreibungen, digitalisierte Fotografien, Akten, Grabungspläne, -skizzen, und -funde, historische Karten, Gebäudepläne, Schwarz-Weiß-Fotografien oder Dias ergänzt. In den darauffolgenden Jahren entstand so eine noch recht überschaubare Datenbank mit einigen tausend Einträgen und 33.000 eingepflegten, digitalisierten Objekten. Dieser Gesamtdatensatz umfasste im Jahr 2011 übersichtliche 500 Gigabyte Speicherplatz.

Die Bedürfnisse von Archivar*innen oder Historiker*innen spielten in dieser ersten Entwicklungsphase noch eine sehr untergeordnete Rolle, sodass die digitale Langzeitarchivierung der in der Datenbank gespeicherten Daten oder die Archivierung der Datenbank als Ganzes noch nicht angedacht wurde.

Dass die Datenbank insbesondere Objekte in den Fokus stellte, wurde durch die in der zweiten Projektphase (siehe nachstehend) einbezogenen Archivare und Historiker als ungewöhnlich empfunden, wurde aber durchaus als Chance für die Weiterentwicklung des Projekts wahrgenommen. Hierdurch könne auch ein in der Forschung zu beobachtender *Shift* weg von reinen Inhalten, hin zu Kulturobjekten und deren sinnlicher Erfassung einbezogen werden.¹¹

4 Entwicklungsstufe 2 (2011–2021): eine offene Austauschplattform

In dieser ersten Projektphase wurden neben der Erfassung des Kulturerbes zwei weitere klare Ziele für die Weiterentwicklung der Datenbank formuliert. Die erfassten Daten sollten unbedingt ebenfalls der Bevölkerung zugänglich gemacht werden, um diese für ihr Kulturerbe zu sensibilisieren. Daneben sollte die Bevölkerung in die inhaltliche Gestaltung der Datenbank einbezogen werden. Hierdurch würden Geschichts- und Kulturinteressierte die Bestandsbildung, die im Rahmen des Projekts vollzogen wurde, mitgestalten können. Gleichzeitig sollte es ihnen ermöglicht werden, über unterschiedliche Medien und Kanäle das Kulturerbe Ostbelgiens (neu) zu entdecken.

4.1 Die Öffnung der Datenbank für neue Akteure

Ab 2011 wurde die bisher existierende Datenbank daher in einer zweiten Projektphase weiterentwickelt. Für diese zweite

8 Vgl. Vortragsskript Ministerium der Deutschsprachigen Gemeinschaft/Tatjana Cormann, Digitalisierung Kulturerbe. Eupen 10. Januar 2019.

9 Vorbildcharakter hatten dabei verschiedene europäische Projekte wie etwa das Informationssystem KuLaDig des Landschaftsverbandes Rheinland (LVR), das (digitale) Inventar für Denkmalkultur der Hauptstadtregion Brüssel oder das KIK-IRPA-Portal des Königlichen Instituts für künstlerisches Kulturgut in Brüssel.

10 Vgl. Vortragsskript Ministerium der Deutschsprachigen Gemeinschaft, 2. Museumskonferenz der Deutschsprachigen Gemeinschaft. Eupen 9. Mai 2016, Folie 15.

11 Exemplarisch wurde hierbei ein Forschungsprojekt der Universität Groningen betrachtet, das sich explizit mit medialen Objekten beschäftigt. Vgl. Annie van den Oever, André Rosendaal, Bernd Warnders, Media Heritage, Final Report, Groningen 2016, online unter: http://media-heritage.org/images/downloads/MediaHeritage_FinalReport.pdf (abgerufen am: 14.07.2023); Ministerium der Deutschsprachigen Gemeinschaft, Protokoll der Arbeitssitzung zum Kulturerbe-archiv und -portal, 21. Januar 2016.

Entwicklungsphase strebte das Ministerium der Deutschsprachigen Gemeinschaft eine *public-public partnership* mit Kulturakteuren aus Ostbelgien an.

Erster Ansprechpartner war hierbei zunächst das Zentrum für Regionalgeschichte in der Deutschsprachigen Gemeinschaft (bis 2017) bzw. das Zentrum für Ostbelgische Geschichte (ZOG), das ein regionalhistorisches Forschungszentrum und eine Vernetzungsplattform für die historisch interessierten Vereinigungen in der Region darstellt. Es gestaltet seit 2011 als durch die Deutschsprachige Gemeinschaft geförderte Vereinigung ohne Gewinnerzielungsabsichten verschiedene Vorhaben im Bereich der Erforschung und Vermittlung von historischem Wissen¹² und verwirklichte in den vergangenen zehn Jahren zahlreiche Projekte im Bereich der Digitalisierung von Quellen. Daneben ist es insbesondere in der historischen Öffentlichkeitsarbeit und der Sicherung von (physischen) Quellen aktiv.

Der zweite Ansprechpartner war das Staatsarchiv in Eupen, das als föderale (Bundes-)Einrichtung des Königreichs Belgien der Deutschsprachigen Gemeinschaft zwar nicht untersteht¹³, aber über wichtige Fachkompetenz im Bereich des (digitalen) Archivwesens verfügt und über zahlreiche Bild-, Karten- und Aktenbestände die Datenbank erheblich erweitern könnte.

Hieraus entstand im Projekt eine neue Rollenverteilung. Geleitet wurde das Projekt weiterhin durch den Fachbereich Kultur des Ministeriums, wobei sich die Deutschsprachige Gemeinschaft zu zahlreichen Dienstleistungen verpflichtete: Sie stellt die Datenbank den neuen Mitgliedern zur kostenlosen Nutzung zur Verfügung, ist für die Weiterentwicklung der Software zuständig, sichert die kostenlose Archivierung der eingespeisten Inhalte zu und begleitet die Projektteilnehmer bei der Einpflege ihrer Daten.¹⁴

Das Zentrum für Ostbelgische Geschichte unterstützte das Ministerium der Deutschsprachigen Gemeinschaft bei dieser zweiten Phase insbesondere bei der Weiterentwicklung der Plattform und bei der Anwerbung sowie Begleitung neuer Projektteilnehmer (cf. infra). Daneben speist es selber Bestände in die Datenbank ein.

Die Rolle des Staatsarchivs in Eupen war in dieser zweiten Phase eher beratender Natur: Es half insbesondere bei der Erweiterung der Metadatenerfassung und war anschließend mit der Einspeisung eigener Quellen beschäftigt.

Das Ziel dieser Erweiterung bzw. der zweiten Projektphase bestand darin, zunächst den Teilnehmerkreis des Projekts zu vergrößern, um somit einen größeren Nutzen aus der bestehenden Plattform zu ziehen. Hierüber sollte natürlich die Erfassung des Kulturerbes, die Vernetzung der im Kulturbereich tätigen Akteure und letztlich ein Bewusstsein für das eigene kulturelle Erbe der deutschsprachigen Minderheit geschaffen werden. Daneben sollte die Sicherung bisher auf regionaler Ebene noch wenig beachteter Quellen wie Filmmaterial, Tonquellen, Oral-History-Interviews oder anderer Multimedia-Quellen stimuliert werden. Ein weiteres Argument war dementsprechend, dass Ostbelgien über das Internet für die historische Forschung zugänglicher gemacht werden sollte. Dies nicht nur für regionale Geschichtsforscher, die historische Arbeit in Schulen, Geschichtsvereine und gesellschaftliche Initiativen, sondern mit Blick auf die Grenzlage des Gebietes auch für ein dialogisches Erinnern und eine gemeinsame Geschichtsarbeit mit den Nachbarregionen sowie anderen europäischen Grenz- und Minderheitenregionen.

In der Folge wurden daher insbesondere durch das Zentrum für Ostbelgische Geschichte Kontakte mit weiteren Partnern geknüpft, die aber keine gleichberechtigte Rolle im Projekt einnehmen. Sie werden durch eine Steuerrungsgruppe, die aus Vertreter/innen der drei genannten Partner besteht, betreut und stellen sogenannte Untermantanten im Projekt dar: Beispiele für diese Untermantanten sind etwa die strukturell geförderten Museen und Geschichtsvereine der Region, ehrenamtlich arbeitende Geschichtsgruppen, Privatpersonen, das Parlament der Deutschsprachigen Gemeinschaft, Vereinigungen und Interessenvertretungen (etwa der Bauernbund oder Musikverband Födekam) und andere Anwender mit historisch oder kulturell relevanten Beständen oder Archiven.¹⁵

12 Vgl. hierzu die Statuten der Vereinigung in Anhang des Belgischen Staatsblattes vom 4. Dezember 2014.

13 Über die Geschichte und Kompetenzen des Staatsarchivs in Eupen: Peter Quadflieg, Die Bestände des Staatsarchivs in Eupen. 3., vollständig überarbeitete und erweiterte Auflage, Brüssel 2019, S. 19–36.

14 Vortragsskript Ministerium der Deutschsprachigen Gemeinschaft, 2. Museumskonferenz der Deutschsprachigen Gemeinschaft. Eupen 9. Mai 2016, Folie 31.

15 Vgl. Vortragsskript Ministerium der Deutschsprachigen Gemeinschaft, 2. Museumskonferenz der Deutschsprachigen Gemeinschaft. Eupen 9. Mai 2016, Folie 16.



Abb. 2: Schematische Darstellung der Projekthierarchie der Kulturerbedatenbank der Deutschsprachigen Gemeinschaft. Eigene Darstellung C0

Durch die Kooperation mit diesen Partnern werden ihre jeweiligen Bestände besser erschlossen und der Öffentlichkeit zugänglich gemacht. Hierbei erfolgt die Dateneingabe in die *Kulturerbedatenbank* dezentral und wird durchaus unterschiedlich durch die einzelnen Mandanten gehandhabt. Während einige von ihnen Arbeitsgruppen eingerichtet haben und regelmäßig Daten einpflegen, fand dieser Prozess bei anderen in punktuellen, abgeschlossenen Projekten statt.

Damit die Mandanten mit der Kulturerbedatenbank vertraut gemacht werden können, finden regelmäßig Workshops und Schulungen mit den Projektpartnern statt. Dies ist eine Grundvoraussetzung für eine Teilnahme am Projekt. Bei diesen Schulungen wird das Hauptaugenmerk insbesondere auf drei Elemente gelegt: Die Teilnehmer erhalten eine Einführung in die Funktionsweise der Datenbank. Daneben wird ihnen das praktische Vorgehen bei der Digitalisierung des Archivguts einschließlich der für ihre Bestände relevanten (Meta-)Informationen sowie den einschlägigen Qualitätsstandards in Bezug auf die Digitalisate sowie die Beschreibung und Erfassung von Metadaten erklärt. Hierbei wird den teilnehmenden Mandanten teilweise entsprechende Hard- und Software zur Verfügung gestellt. Ein letzter wichtiger Aspekt bei diesen Schulungen ist eine Einführung in rechtliche Fragen mit Blick auf die Speicherung und Zugänglichmachung von Archivgut, wobei die Teilnehmer/innen die Grundlagen des Urheber- und Verwertungsrechts kennenlernen sowie mit den Standards des *Creative Commons*-Lizenzsystems, mit dem in der *Kulturerbedatenbank* gearbeitet wird, vertraut gemacht werden.¹⁶

Hierbei zeigt die Erfahrung, dass die Funktionsweisen der Kulturerbedatenbank und des dazugehörigen *Interfaces* sehr gut durch die Nutzer/innen angenommen werden. Hierfür musste in dieser zweiten Entwicklungsphase ein benutzerfreundliches *Interface* entwickelt werden, das auch weniger versierten Personen die Möglichkeit bietet, an dem Projekt teilzunehmen und eigene Inhalte in das Projekt einzupflegen. Zentral stand dabei der Wille, eine möglichst einfache und allgemeinverständliche Benutzeroberfläche zu entwickeln, die für die Untermantanten nur über ein Mindestmaß an Funktionalitäten verfügt.

Auch die Digitalisierung von Archivgut und die hiermit verbundenen technischen Fragen stellen keine größeren Hindernisse für eine Teilnahme am Projekt dar, da den Teilnehmer/innen Ansprechpartner an die Hand gegeben werden und im Intranet der *Kulturerbedatenbank* die Möglichkeit des Versendens von Fragen per Direktnachricht besteht. Neben den Workshops, Schulungen und Vorträgen, die als Einführung stattfinden, wurden den Nutzer/innen auch detaillierte und verständliche Handreichungen zur Verfügung gestellt. Daneben ist die *Kulturerbedatenbank* so aufgebaut, dass neben einer nachgeschalteten manuellen Kontrolle durch die betreuenden Mitarbeiter/innen des Zentrums für Ostbelgische Geschichte und des Ministeriums der Deutschsprachigen Gemeinschaft auch immer eine vorgeschaltete softwarebasierte Kontrolle der eingepflegten Daten und Metadaten stattfindet.¹⁷

16 Vgl. hierzu <https://creativecommons.org/about/cclicenses/> (abgerufen am: 14.07.2023). Gerade die letztgenannte Einführung in rechtliche Fragen erweist sich allerdings immer wieder als relativ schwierig. Natürlich legt die Projektleitung – auch aus eigenem Interesse – viel Wert auf die Einhaltung von Autorenrechten, Nutzungsrechten und Persönlichkeitsrechten. Diese spielen natürlich für eine Archivierung im Rahmen der Kulturerbedatenbank eine eher untergeordnete Rolle, werden aber dann zum Problem, wenn Bilder veröffentlicht werden. Die Erfahrung der letzten Jahre hat gezeigt, dass eher wenig Feingefühl unter den Nutzer/innen für diese rechtlichen Fragen besteht.

17 Es dürfen etwa nur je digitalisiertem Datenträgertyp bestimmte Formate eingespeist werden. Ein weiteres Beispiel für eine softwarebasierte Qualitätskontrolle ist das Abfragen und die Kontrolle von Pflichtfeldern bei der Einspeisung von Metadaten.

4.2 Verknüpfung und Vernetzung mit und im Projekt

Zur bestmöglichen Vernetzung und gleichzeitigen Öffnung der *Kulturerbedatenbank* wurde eine Mehrebenenstruktur für die Datenbank entwickelt, die es auf der ersten Stufe den Projektteilnehmer/innen ermöglicht, Dokumente für die Öffentlichkeit unzugänglich zu speichern, im Open Intranet – auf der zweiten Stufe – mit den Projektteilnehmern zu teilen und letztlich auf der dritten Stufe für das Internet auf mit der Datenbank verknüpften Portalen frei zugänglich zu machen.

Diese Herangehensweise entsprang der Beobachtung und dem Vergleich mit anderen Digitalisierungs-, Archivierungs- und Multimediaprojekten, die sich insbesondere auf die Archivierung und die Digitalisierung von analogem Archivgut konzentrierten. In den Fokus dieser zweiten Projektphase rückte daher insbesondere die spätere Nutzung der eingespeisten Daten, der Mehrwert für die Endnutzung des Projektes und die Vernetzung der Teilnehmer im Projekt.¹⁸

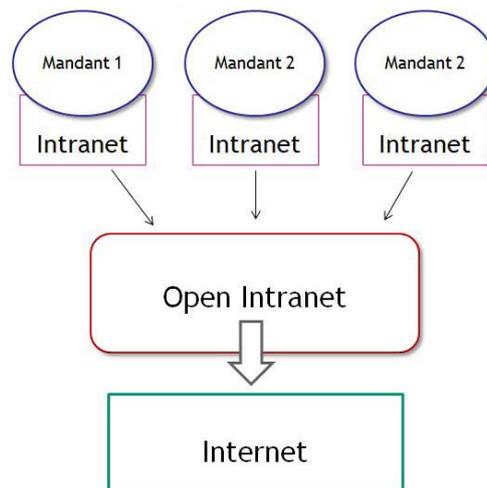


Abb. 3: Mehrebenenfreigabe im Kulturerbearchiv der Deutschsprachigen Gemeinschaft, Ministerium der Deutschsprachigen Gemeinschaft (CC BY-NC-ND 4.0)

Hierbei dient die erste Stufe der Mehrebenenstruktur zur Verwaltung der eigenen, eingespeisten Daten und von dort aus können auch Archivalien in die *Kulturerbedatenbank* eingepflegt werden. Auf dieser Ebene existiert ebenfalls ein persönlicher Bereich, in dem mit den anderen Projektteilnehmern über eine Nachrichtenfunktion in Kontakt getreten werden kann. Hier erfolgt ebenfalls die Freigabe (oder nicht) von Objekten für das Open Intranet.

Das Open Intranet ist ein Teilbereich, in dem die Projektteilnehmer/innen Daten untereinander teilen können und der ebenfalls für temporäre Forscher/innen oder Besucher/innen zugänglich gemacht werden kann. Dieser Bereich ermöglicht es den Teilnehmer/innen nicht nur, die Sammlungen der anderen Einpflegenden zu entdecken, sondern eröffnet auch die Option, die Metadaten der anderen Sammlungen mit neuen Informationen anzureichern.

Dabei achtet die Projektleitung darauf, dass die einzelnen Teilnehmer/innen genügend Daten auch für das Open Intranet freigeben. Hierdurch soll einer unverhältnismäßigen Nutzung einer aus öffentlicher Hand finanzierten Plattform von privater Seite entgegengewirkt werden. Dabei ist der Umfang dieser für das Intranet freizugebenden Daten davon abhängig, um welche Archivalien es sich handelt: Urheberrechtlich geschütztes Material oder Archivalien, bei denen die Verwertungsrechte nicht geklärt sind, müssen nicht zwangsläufig für das Open Intranet freigegeben werden.

In letzter Instanz besteht die Möglichkeit, die Daten für das Internet freizugeben. Für die Veröffentlichung etwa im Rahmen einer Website müssen natürlich (teils angepasste) Schnittstellen vorgesehen werden. Einen ersten Schritt in diese Richtung machte das Ministerium der Deutschsprachigen Gemeinschaft bereits im Jahr 2011, als es sein „Kulturerbeportal“ veröffentlichte.¹⁹ Über dieses konnten die ersten, oben erwähnten Datensätze zum Denkmal- und Landschaftsschutz durchsucht werden.

Daneben verknüpften das Töpfereimuseum in der Ortschaft Raeren (<https://www.toepfereimuseum.org/>) sowie das Zentrum für Ostbelgische Geschichte (<https://www.geschichte.be/>) ihre Websites mit der Kulturerbedatenbank und nutzten deren Funktionalitäten auf sehr unterschiedliche Weise.

Neben der Möglichkeit, die *Kulturerbedatenbank* an externe Internetauftritte anzubinden und den Export von Inhalten zu erlauben, wurden weitere Schnittstellen zur Veröffentlichung von Inhalten geschaffen: Jeder Datenbankeintrag

18 Ministerium der Deutschsprachigen Gemeinschaft, Protokoll der Arbeitssitzung zum Kulturerbearchiv und -portal, 21. Januar 2016.

19 Für eine Version aus dem Jahr 2012 des Kulturerbeportals vgl. <https://web.archive.org/web/20120512061035/http://www.dgkulturerbe.be/desktopdefault.aspx> (abgerufen am: 14.07.2023). Eine Datenbankabfrage ist über die im Internet-Archive archivierte Version nicht mehr möglich.

verfügt beispielsweise über die Möglichkeit, einen QR-Code zu generieren, über den die Einträge beispielsweise bei Themenwanderungen abgerufen werden können. Gleichzeitig wurden diese QR-Codes in die Denkmalschutzplaketten der Deutschsprachigen Gemeinschaft integriert, sodass vor jedem denkmalgeschützten Gebäude über ein mobiles Endgerät ein Datenbankabruf erfolgen kann.

4.3 Neue Schwerpunkte durch neue Nutzerkreise

Dabei musste die Datenbank zunächst auf die Bedürfnisse neuer Mitglieder ausgelegt werden. War das Kulturerbearchiv bis in die 2010er-Jahre vor allem auf die Erfassung des architektonischen bzw. archäologischen Kulturgutes ausgelegt, wurde nunmehr auch mit den Bedürfnissen von Archiven, Forscher/innen, Geschichtsvereinen, Museen und Sammler/innen bzw. interessierten Dritten Rechnung getragen. Dies führte bei der Neuorientierung zwar immer wieder zu Verständigungsschwierigkeiten zwischen Angehörigen unterschiedlicher Kultursparten und den IT-Dienstleistern, konnte aber durch die ursprüngliche Veranlagung der relationalen Objektdatenbank recht einfach aufgefangen werden:

In der Praxis bedeutete dies für die Struktur der Datenbank, dass sich durch das Aufeinandertreffen verschiedener Interessen insbesondere die Rolle der bis dato zentral stehenden Objekte wandelte. Im Denken der zweiten Projektphase wurden die Objekte insbesondere aus der Perspektive der teilnehmenden Archivar/innen und Historiker/innen zu den tiefsten Beschreibungseinheiten bzw. den zentralen Verzeichnungseinheiten.

Diese Objekte können nach der Erstellung in der Datenbank ebenfalls Sammlungen und Beständen zugeordnet werden.²⁰ Gleichzeitig erlauben sie das Hinzufügen von untergeordneten Verzeichnungselementen, die durch die Archivar/innen des Staatsarchivs als *deelbeschrijving* bezeichnet werden.²¹

Begründet war diese praktische Verlagerung des Stellenwerts der Objekte insbesondere durch die intensivere Arbeit mit neuen Quellen. In den Fokus der zweiten Projektphase rückten die Digitalisierung von Fotografien und die Einspeisung von digitalisierten historischen Zeitungen.

Eine Veränderung erfuhr dementsprechend auch die Erfassung der Metadaten. Bis zur Erweiterung des Projekts um die archivisch bzw. historisch arbeitenden Partner, beschränkte sich die Erhebung der Metadaten auf einige zentrale und häufig frei formulierbare Angaben zu einzelnen Verzeichnungseinheiten. Vor dem Hintergrund eines professionell arbeitenden Personenkreises und einer Vielzahl unterschiedlicher Quellen war das wohl in der ersten Projektphase eine gerechtfertigte Herangehensweise.

Da in der zweiten Projektphase auch die interessierte Öffentlichkeit an der Kulturerbedatenbank beteiligt wurde, konnte die Eingabe von Metadaten nicht mehr offen gestaltet werden. Je eingespeister Quellenart wurde daher ein eigenes Datenmodell zur Erfassung der Metadaten entwickelt. Für Fotografien dienten die der *Europeana* zugrundeliegenden Datenmodellen als Vorbild.²² Für schriftliche Archivalien oder Akten richtet sich die Erfassung der Metadaten nach den im Belgischen Staatsarchiv angewendeten Verzeichnungsnormen, d. h. ISAD(G).²³ Für die Verschlagwortung der eingepflegten Inhalte wird ebenfalls auf bereits Bestehendes zurückgegriffen: In diesem Zusammenhang arbeitet das Projekt beispielsweise mit dem „Wortnetz Kultur“, dem Thesaurus des Landschaftsverbands Rheinland (LVR).

Dabei existiert ebenfalls die Möglichkeit, softwarebasiert Metadaten hinzuzufügen. Der Nutzer hat etwa die Möglichkeit, Filme oder Tondateien über die Software *Abby Finereader* transkribieren zu lassen. Hierdurch sollen beispielsweise die im Parlament der Deutschsprachigen Gemeinschaft aufgenommenen Wortbeiträge erfasst und besser zugänglich gemacht werden. Ebenfalls besteht die Möglichkeit, Textdateien über OCR-Erfassung auslesen zu lassen. Hierdurch können die digitalisierten Quellen mit neuen Metadaten angereichert werden, was zweifelsohne eine bessere Nutzung ermöglicht.

5 Fazit: Eine ausbaufähige (?) Austauschplattform

Nachdem diese beiden Entwicklungsschritte als abgeschlossen gelten können, muss in der Gesamtschau ein recht positives Fazit mit Blick auf die Gestaltung der *Kulturerbedatenbank* gezogen werden. In den zwei Projektphasen konnten bereits eine halbe Million unterschiedliche Dokumente in die Datenbank eingepflegt, erfasst und gesichert werden. Diese stammen vielfach aus Privatarchiven und sind nunmehr einer breiteren Öffentlichkeit über unterschiedliche Schnittstellen zur Verfügung gestellt worden.

Das Projekt wird durch den Kreis der potenziell Interessierten sehr gut angenommen. Ein wichtiger Pfeiler ist dabei die konstante Begleitarbeit, die durch das Ministerium der Deutschsprachigen Gemeinschaft und das Zentrum für

20 Ministerium der Deutschsprachigen Gemeinschaft, Protokoll der Arbeitssitzung zum Kulturerbearchiv und -portal, 21. Januar 2016.

21 Herman Coppens, *Directives relatives au contenu et à la forme d'un inventaire d'archives*, Brüssel 2014, S. 68.

22 Vgl. Vortragsskript Ministerium der Deutschsprachigen Gemeinschaft, 2. Museumskonferenz der Deutschsprachigen Gemeinschaft. Eupen 9. Mai 2016, Folie 20.

23 Je nach Bedarf können der Datenbank auch andere „Objekte“ hinzugefügt werden. Angedacht ist etwa die Erweiterung der Datenbank um die Möglichkeit, biografische Angaben einzupflegen. Prinzipiell würden Personen zu den zentralen Verzeichnungseinheiten werden.

Ostbelgische Geschichte geleistet wird. Dabei erlaubt das Projekt, das Verknüpfen von unterschiedlichen Nutzer/innen und deren Beständen sowie die Zurverfügungstellung einer kostenintensiven Infrastruktur, die sonst lediglich staatlichen Archiven oder Verwaltungen zugänglich wäre. Hierdurch entstand bereits jetzt eine aus regionalhistorischer Perspektive umfangreiche digitale Datenbank, die aus allen Teilen des Gebiets deutscher Sprache Belgiens Quellen beinhaltet und diese häufig physisch verstreute Überlieferung dementsprechend verknüpft. Besonders hervorzuheben ist aus Perspektive des Historikers vor diesem Hintergrund, dass die Datenbank nicht den Bestand eines einzelnen staatlichen Archivs abbildet, sondern dezentral aufgebaut wurde und hieraus ein vielfältiger Quellenbestand aus der „bottom-up“ Perspektive entstand. Bemerkenswert waren dabei einige Nebeneffekte des Projektes: Während man zweifelsohne mit einer Neu-Inwertsetzung des Kulturerbes gerechnet hatte, war auch aufseiten der Projektteilnehmer/innen eine Neufokussierung auf bisher lediglich physisch vorhandene Bestände zu beobachten: Die Digitalisierung machte dabei die Originale keinesfalls überflüssig, sie erhöhte eher ihre Sichtbarkeit und die Sensibilität für deren kulturgeschichtlichen Wert. Sie weckte erneut Interesse an der Sache und verstärkte den Wunsch, die Originale selbst neu zu entdecken.

Zweifelsohne wohnt der Kulturerbedatenbank aber noch enormes Entwicklungspotenzial inne. Der Datenbank liegen beispielsweise keinerlei Standards zur sogenannten digitalen Langzeitarchivierung zugrunde. Es bestehen zwar Redundanzen, sodass die Dateien und Metadaten auf unterschiedlichen Servern vorliegen, aber hierbei werden kaum Standards für die digitale Langzeitarchivierung erfüllt. Für die langfristige Sicherung der Datenbank zeichnen sich daher mehrere Optionen ab. Angedacht wird etwa eine verstärkte Kooperation mit dem flämischen Institut für digitale Archivierung *Meemoo*, das ein ausgewiesenes Expertisezentrum für die Digitalisierung und Archivierung von audiovisuellen Quellen ist.²⁴ Hierbei würde nicht die existierende relationale Datenbank archiviert werden, sondern lediglich die Datensätze und Metadaten durch *Meemoo* zur Einspeisung in eine neue Infrastruktur und zur Sicherung übernommen werden.

Ein weiterer wichtiger Zukunftsaspekt betrifft die Komplexität der Datenbank. Diese ist jetzt schon auf sehr unterschiedliche Bedürfnisse, Formate und Schnittstellen ausgelegt, wodurch sie einen gewissen Komplexitätsgrad erreicht hat. Um möglichst für eine breite Öffentlichkeit zugänglich zu bleiben und eine effiziente Nutzung der Datenbank zu gewährleisten, muss dieser Aspekt in Zukunft im Auge behalten werden. Hierbei hat die Erfahrung mit verschiedenen Datenbanken gezeigt, dass in Hinblick auf die Benutzerfreundlichkeit und Geschwindigkeit „weniger“ häufig „mehr“ ist.²⁵

In einer dritten Projektphase könnte ebenfalls angedacht werden, bereits vorhandene Visionen in die Tat umzusetzen. Obwohl die Metadatenerfassung auf eine Teilnahme am Europeana-Projekt ausgelegt ist, fand eine Verknüpfung mit dem Portal bisher noch nicht statt.

Ein letzter Wermutstropfen sei an dieser Stelle ebenfalls angesprochen. Obwohl viele enthusiastische und geschichtsbegeisterte Arbeitsgruppen für eine Teilnahme am Projekt begeistert werden konnten, scheiterte eine intensivere Beteiligung häufig an einem nicht zu vernachlässigendem Aspekt: Der Geschwindigkeit des Internets im ländlichen Raum ...

24 Vgl. hierzu <https://meemoo.be/nl> (abgerufen am: 14.07.2023). Meemoo beschäftigt sich bis dato insbesondere mit der digitalen Archivierung von audiovisuellen Quellen, wobei es mit über ganz Flandern verteilten Kontentpartnern zusammenarbeitet.

25 Vgl. hierzu auch das Beispiel des Archivs der Stadt Nürnberg Annette Birkenholz.

Durchsuch- & Extrahierbarkeit als signifikante Eigenschaften von PDF/A

Claire Röthlisberger-Jourdan, Georg Büchler, KOST

1 Vorbemerkung

Die hier präsentierte Problematik wurde von der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST, <https://kost-ceco.ch/>) 2020 in einer ausführlichen Studie untersucht.¹ Um unnötige Doppelungen zu vermeiden, geht dieser Tagungsbeitrag nur soweit auf die technischen Grundlagen ein, als sie für das Verständnis unverzichtbar sind, und verweist für Details auf die erwähnte Studie. Er vertieft stattdessen archivtheoretische Aspekte, welche die Problematik aufwirft.

2 Problem

Anlass zu den hier zusammengefassten Untersuchungen waren als valide erkannte PDF/A-Dokumente, aus denen die Suche und der Export von Text nicht oder nur fehlerhaft möglich ist (Abbildung 1). Da PDF/A die Einbettung von Schriftarten fordert, ist dieser Befund auf den ersten Blick kontraintuitiv. Um ihn zu verstehen, ist es notwendig, die verschiedenen Grundsätze der Schriftarteneinbettung in PDF/A zu analysieren.²

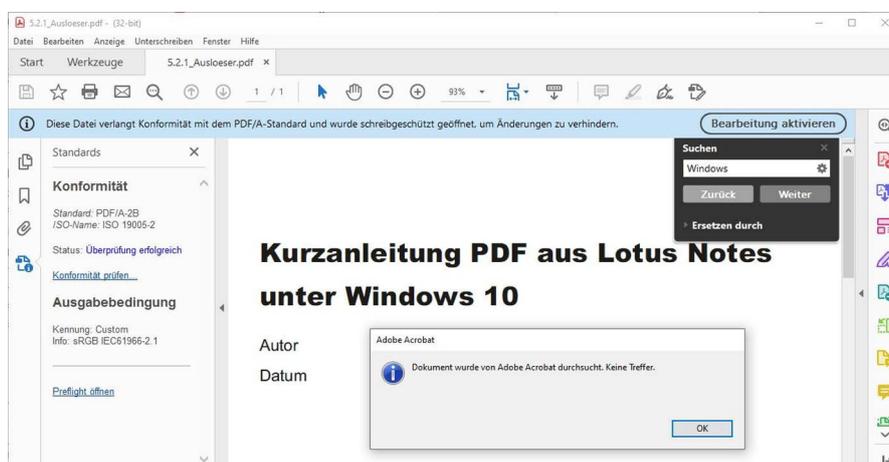


Abb. 1: Suchergebnis nach dem Wort «Windows». KOST

Damit eine PDF/A-Datei des Konformitätsniveaus B (PDF/A-1b, PDF/A-2b, PDF/A-3b) valide ist, müssen im Minimum die Konturen der verwendeten Zeichen eingebettet sein. Das ist in den vorliegenden Dokumenten der Fall. Für die Texterkennung und -extraktion ist dies aber nicht hinreichend, sondern dafür wird für jedes Zeichen zusätzlich die Zuordnung zum korrekten Unicode-Zeichen benötigt, welche direkt oder über eine eingebettete CMap „ToUnicode“ realisiert werden kann. Wenn diese fehlt (Abbildung 2), wird bei der Extraktion je nach Anwendung entweder „Unbekanntes Zeichen“ angegeben oder der Standardwert für die Position des Zeichens innerhalb der eingebetteten Schrift angenommen, welcher richtig oder falsch sein kann.



Abb. 2: Extraktionsergebnisse des Wortes „Windows“. KOST

- 1 KOST, PDF mit mangelhaftem Font: Text ist nicht durchsuch- und extrahierbar, Bern 2020, online unter: <https://kost-ceco.ch/cms/pdf-nicht-durchsuchbar.html> (abgerufen am: 14.07.2023).
- 2 PDF/A ist als ISO-Standard in inzwischen vier Versionen publiziert, zuerst 2005: ISO 19005-1:2005, Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1), Genf 2005, online unter: <https://www.iso.org/standard/38920.html> (abgerufen am: 14.07.2023). Für eine ausführliche Diskussion der PDF/A-Versionen 1 bis 3 siehe KOST, Katalog archiverischer Dateiformate KaD, Version 6.1, Bern 2020, online unter: <https://kost-ceco.ch/cms/pdf-a-1.html>, <https://kost-ceco.ch/cms/pdf-a-2.html> und <https://kost-ceco.ch/cms/pdf-a-3.html> (abgerufen am: 14.07.2023).

3 Analyse

Zum Zeitpunkt dieser Entdeckung existierte keine Möglichkeit, rasch und effizient zu eruieren, ob eine PDF-Datei vom beschriebenen Problem betroffen ist. Um Klarheit über das Ausmaß des Problems in ihren Trägerarchiven zu erhalten, hat die KOST die Firma PDF Tools AG beauftragt, ein entsprechendes Zusatzmodul für den „3-Heights TM PDF Validator“ zu entwickeln, welcher auch in KOST-Val, dem Formatvalidator der KOST,³ integriert ist. Damit konnten PDF/A-Bestände der Ausprägungen PDF/A-1b und PDF/A-2b aus den Staatsarchiven St. Gallen, Basel-Stadt und Uri, aus dem Schweizerischen Bundesarchiv sowie Testdaten der KOST-Geschäftsstelle analysiert werden. Im Schnitt waren rund 20% aller Dateien betroffen.

Nicht alle so erkannten Fehler sind gleich schwerwiegend. Als geringfügig bezeichnen wir beispielsweise Fehler in Symbolschriften oder bei bestimmten, bekannten Zeichen wie Abständen oder Aufzählungszeichen. Für die detaillierte Argumentation dazu verweisen wir auf die KOST-Studie.⁴ Eine zusätzliche Konfiguration im Formatvalidator KOST-Val erlaubt es, solche geringfügigen Fehler zu ignorieren, um die Identifikation und Fehlerbehebung auf die wirklich problematischen Fälle zu begrenzen. Damit sind noch rund 1% aller Dateien betroffen.

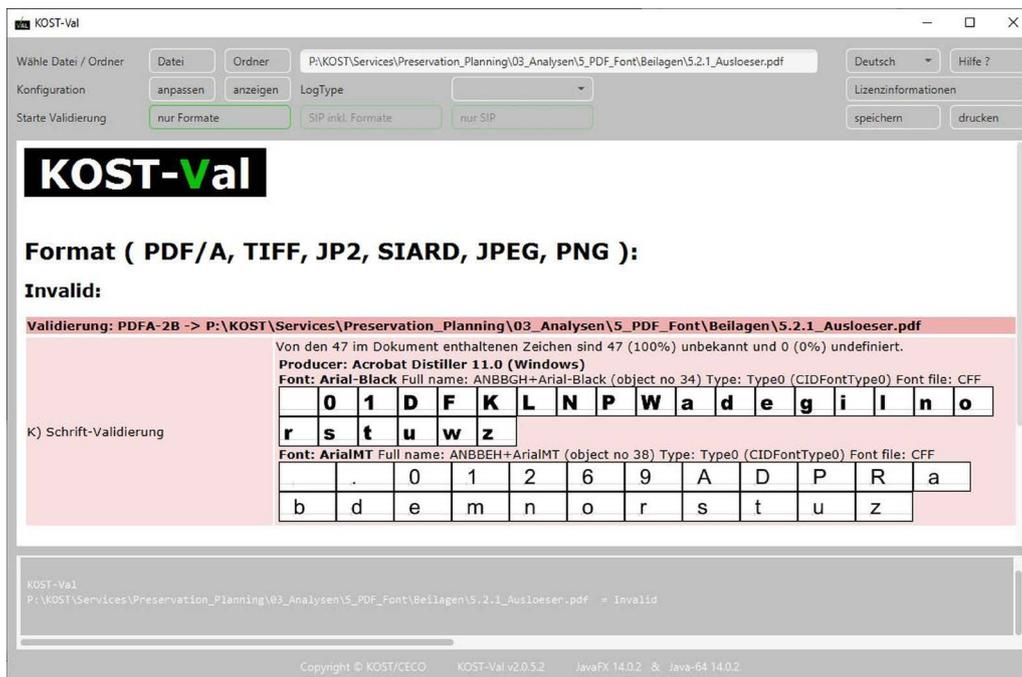


Abb. 3: Validierungsergebnis mit KOST-Val. KOST

Im Minimum sollten solche Dateien im Archiv entsprechend gekennzeichnet werden. Zusätzlich kann eine Mängelbehebung angezeigt sein, wobei die fehlenden Abbildungen auf Unicode ergänzt werden. Bei einer nicht korrekt eingebetteten Schrift genügt dazu in der Regel eine Konvertierung nach PDF/A-2u, wenn die Schrift auf dem Rechner installiert ist. Wenn nur die Kontur der Zeichen eingebettet ist, können diese von einer OCR-Engine ausgewertet und die entsprechenden Unicode-Werte in der Schrifttabelle ergänzt werden.

4 Diskussion

Im Folgenden sollen zwei Lehren aus dieser Untersuchung mit Blick auf den archivischen Umgang mit PDF/A entwickelt werden.

4.1 Sich entwickelnde signifikante Eigenschaften

Die nestor-Arbeitsgruppe Bestandserhaltung hat in ihrem Leitfaden zur digitalen Bestandserhaltung in Ergänzung und Abweichung zum OAIS-Terminus der „Designated Community“⁵ das Konzept der Nutzungsziele von Informationen

3 KOST, KOST-Val, online unter: <https://kost-ceco.ch/cms/kost-val.html> (zuletzt abgerufen am: 14.07.2023).

4 KOST, wie Anm. 1, S. 4f.

5 Siehe ISO 14721:2012, Space data and information transfer systems — Open archival information system (OAIS) — Reference model, Genf 2012, S. 1–11 und öfter, online unter: <https://www.iso.org/standard/57284.html> (abgerufen am: 14.07.2023).

eingeführt und argumentiert, dass sich aus diesen in sinnvoller Weise Erhaltungsgruppen bilden und deren signifikante Eigenschaften (Significant Properties) ableiten lassen.⁶ Dieses Konzept stellt sich als hilfreich heraus für die Analyse des hier zugrundeliegenden Problems. Bei der Entwicklung von PDF/A-1 in den Jahren vor 2005 wurde als implizites Nutzungsziel die menschliche Lesbarkeit des Dokuments identifiziert und folglich die einwandfreie und systemunabhängige visuelle Wiedergabe als signifikante Eigenschaft spezifiziert.⁷ Seither haben Verfahren wie Volltextsuche, Textextraktion und Text Mining in der archivischen Forschung und Praxis stark an Bedeutung gewonnen. Diese implizieren ein neues Nutzungsziel, nämlich die maschinelle Verarbeitung des in Dokumenten enthaltenen Textes, und folglich zwei neue signifikante Eigenschaften, die Durchsuchbarkeit sowie die Extrahierbarkeit des Textes. Diese Eigenschaften müssen zu einer Anpassung der Anforderungen an Dateiformate zur Archivierung textbasierter Unterlagen führen. Folgerichtig empfiehlt die KOST, bei der Konversion zu PDF/A die Ausprägung PDF/A-2u zu verwenden.⁸ Das Konformitätslevel U verlangt zusätzlich zu den Anforderungen des Levels B die eindeutige Abbildung der eingebetteten Zeichen auf Unicode, womit das hier zugrundeliegende Problem behoben wird.

4.2 Formatspezifikationen als Minimalvorgaben

Das Problem der fehlerhaften Textextraktion aus PDF/A-Dokumenten des Konformitätslevels B ist auch ein Lehrstück zur genauen Interpretation von Standards. PDF/A macht summa summarum nur minimale, zentrale Vorgaben zu erlaubten, erforderlichen und verbotenen Eigenschaften von konformen Dateien. In der Praxis übererfüllen PDF/A-Dokumente vielfach den Standard. So ist ein Großteil der PDF/A-1b-Dokumente in Archiven problemlos durchsuch- und extrahierbar, weil die eingebetteten Schriftarten korrekt auf Unicode abgebildet sind. Eine Archivierungsstrategie darf sich aber nicht auf eine solche Übererfüllung verlassen. Wenn die Durchsuchbarkeit oder die Extrahierbarkeit des Textes eine signifikante Eigenschaft ist, muss das Archiv das Konformitätsniveau U verlangen.

5 Fazit

Was zunächst wie ein Fehler aussah, nämlich die Unmöglichkeit der Texterkennung und -extraktion aus PDF/A-Dateien, erwies sich bei genauerer Betrachtung als strikte Erfüllung der Vorgaben von PDF/A-1b beziehungsweise PDF/A-2b. Das eigentliche Problem liegt darin begründet, dass sich ändernde Nutzungsziele nicht in die Anforderungen an Archivformate überführt wurden. Archive sind gut beraten, sich solcher (oftmals schleichenden) Änderungen bewusst zu werden und diese explizit auf signifikante Eigenschaften und in der Folge auf Anforderungen an Archivformate abzubilden.

6 nestor-Arbeitsgruppe Digitale Bestandserhaltung, Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung, Version 2.0, nestor-Materialien 15, [o.O.] 2012, S. 24ff., online unter: <http://nbn-resolving.de/urn:nbn:de:0008-2012092400> (abgerufen am: 14.07.2023). Das Konzept der Significant Properties wurde 1999 entwickelt vom Cedars Projekt, online unter: <http://www.ukoln.ac.uk/metadata/cedars/> (abgerufen am: 14.07.2023), und seither intensiv rezipiert.

7 Dies gilt als Minimalanforderung für alle Konformitätsstufen. Die Konformitätsstufen A und (ab Version 2) U gehen darüber hinaus.

8 KOST, Katalog archivischer Dateiformate KaD, Version 6.1, PDF/A-2, Bern 2020, online unter: <https://kost-ceco.ch/cms/pdf-a-2.html> (abgerufen am: 14.07.2023).

Digitalisierung im öffentlichen Hochbau

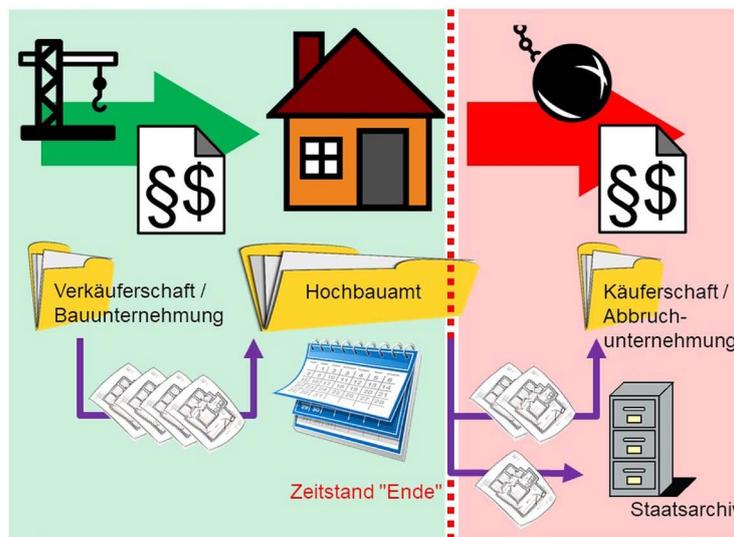
Eine Chance auch für die Archivierung

Flurina Camenisch, Staatsarchiv Graubünden, und Bernhard Stüssi, Staatsarchiv Zürich

1 Die Archivierung von Gebäudedossiers als gordischer Knoten

Die öffentliche Hand ist auch in der Schweiz eine bedeutende Eigentümerin und Betreiberin von Hochbauten. Für die Verwaltung der kantonalen Gebäude ist in Graubünden das Hochbauamt zuständig, mit welchem das Staatsarchiv sehr gut zusammenarbeitet. Die Serie der pro Gebäude angelegten Dossiers dokumentiert die Kernaufgabe des Hochbauamts; entsprechend hoch ist ihr Archivwert. Allerdings brachte in der Vergangenheit die Archivierung von analogen Gebäudedossiers bewährte archivische Konzepte an ihre Grenzen, ja führte sie nahezu ad absurdum. Das seit 2020 beim Hochbauamt geltende digitale Primat erleichtert nicht nur die Arbeit dieser Behörde, sondern bietet auch für die Archivierung große Vorteile.

Die folgende Abbildung zeigt den Lifecycle von (analogen) Gebäudedossiers des Hochbauamts bis 2019.



Bei Erstellung oder Kauf eines Gebäudes durch den Kanton legte das Hochbauamt ein Dossier an, das sich wesentlich aus der Dokumentation der Verkäuferschaft bzw. der am Bau beteiligten Unternehmen speiste.

Solange das Gebäude im Besitz des Kantons war, stellte es für das Hochbauamt ein offenes Geschäft dar. Im Rahmen von Unterhalts- und Umbauarbeiten wurde auf die Pläne, Verträge und technischen Spezifikationen im Gebäudedossier zurückgegriffen. Laufend wurden neue Dokumente in das Dossier gelegt, während obsolete entfernt wurden.

Bei Veräußerung oder Rückbau des Gebäudes wurden der Käuferschaft respektive der Abbruchunternehmung Kopien aus dem Dossier übergeben. Das Hochbauamt schloss das entsprechende Geschäft und lieferte dem Staatsarchiv das Originaldossier in etwas ausgedünnter Form ab. So war sichergestellt, dass im Archiv zu allen Gebäuden, die ehemals im Eigentum des Kantons waren, Unterlagen zu finden sind.

Dieses System war aus archivischer Sicht in verschiedener Hinsicht unzulänglich:

- Das Konzept, wonach ein Geschäftsdossier eröffnet und geschlossen wird, geht davon aus, dass die Bearbeitung des Geschäfts über die Zeit einigermaßen gleichförmig vonstattengeht. Wenn aber ein Gebäudedossier über Generationen offenblieb, was nicht selten vorkam,¹ wurde es durch Veränderungen in Bauwesen, Politik, der Art und Weise der Gebäudebewirtschaftung und den dokumentarischen Vorgaben meist so beeinträchtigt, dass man dabei eher von einem Generationenprojekt als von *einem* Geschäft sprechen kann;
- Der Kanton besitzt eine große Vielfalt an Gebäudetypen (sie reicht vom Kunstmuseum bis zur Fischzucht), deren Dokumentationen inhaltlich unterschiedlichen Anforderungen zu genügen haben und sich deshalb oft erheblich voneinander unterscheiden;
- In der Regel wurde auch bei Bauvorhaben und Handänderungen keine Dokumentation nach einem einheitlichen Muster erstellt, sondern es wurden diejenigen Dokumente aufbewahrt, die im konkreten Fall geschäftsrelevant waren, ohne die längerfristige Nachvollziehbarkeit der „Geschichte“ des jeweiligen Gebäudes besonders zu berücksichtigen;

1 Das 1752 erbaute Regierungsgebäude z. B. nutzte bereits der Rechtsvorgänger des 1803 geschaffenen Kantons Graubünden. Seit 1807 gehört es dem Kanton.

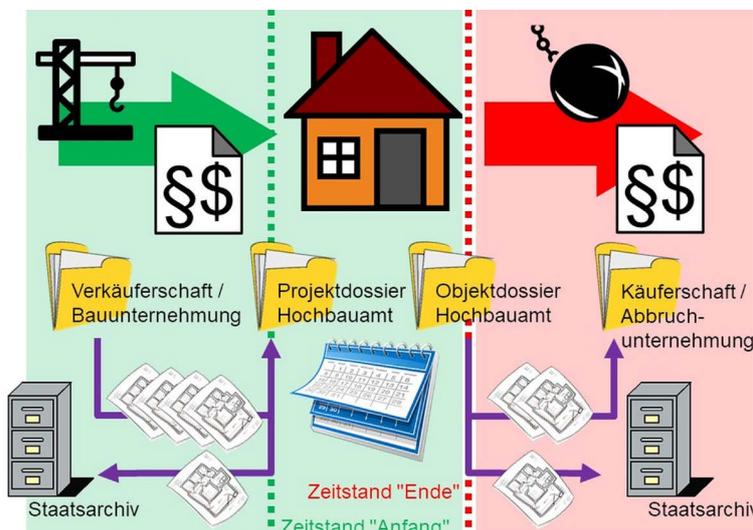
- Im Verlauf ihrer meist Jahrzehnte und nicht selten Jahrhunderte umfassenden Lebensdauer werden Gebäude in der Regel mehrmals renoviert, um- und angebaut. Diese Eingriffe in die Bausubstanz hatten beträchtliche Auswirkungen auf den Inhalt der Gebäudedossiers;
- Die Dokumentation wurde nur zu dem Zeitpunkt archiviert, in dem das Gebäude das kantonale Eigentum verließ. Ältere Zeitstände sind meist nur schwer oder gar nicht mehr rekonstruierbar;
- Wurde ein Gebäude rückgebaut oder veräußert, hatte das Hochbauamt keine Verwendung mehr für das betreffende Dossier; die Aufbewahrungsfrist betrug null Jahre. Es hätte kassiert werden können, hätte das Staatsarchiv nicht den Archivierungsvorbehalt geltend gemacht. Dies wiederum bedeutete, dass das Hochbauamt die Dossiers für das Archiv (von Hand) ausdünnen und bereinigen musste.

Am Staatsarchiv sind zwar zu sämtlichen ehemals kantonalen Hochbauten Dossiers vorhanden,² von einer Nachvollziehbarkeit des „Gebäude-Geschäftsfalls“ kann aber meist nicht die Rede sein.

2 Die Digitalisierung zerschlägt den Knoten

Seit 2020 verwaltet das Hochbauamt die Gebäudedossiers digital. Dieses neue System behebt die genannten Unzulänglichkeiten und bietet sowohl der Fachbehörde als auch dem Staatsarchiv Möglichkeiten, die es im „analogen Zeitalter“ nicht gab.

Die folgende Abbildung zeigt den Lifecycle von (digitalen) Gebäudedossiers des Hochbauamts seit 2020.



Neu wird unterschieden zwischen Projekt- und Objektossier. Für die Bauphase oder die Übernahme eines Gebäudes wird ein Projektossier angelegt. Bei Inbetriebnahme des Gebäudes durch den Kanton – dieser Zeitpunkt ist durch die Genehmigung der Schlussabrechnung genau definiert – wird das Projektossier geschlossen.

Für die Betriebsphase des Gebäudes wird ein neues Objektossier eröffnet, in das die weiterhin benötigten Unterlagen aus dem Projektossier kopiert werden. Im Objektossier bilden die kopierten Unterlagen vollwertige neue Originale.³ Das Objektossier wird während der Betriebsphase wie bisher durch das Hochbauamt bewirtschaftet. Bei Veräußerung oder Rückbau des Gebäudes wird auch das Objektossier geschlossen und dem Staatsarchiv abgeliefert.

Stehen Umbauten oder Renovationen an, entscheidet das Hochbauamt, ob diese im laufenden Betrieb ausgeführt werden, oder ob es sich um Projekte und damit um separate Geschäftsfälle handelt. Die Beurteilung erfolgt aufgrund des Investitionsvolumens, der technischen Komplexität und des Ausmaßes des Eingriffs in die Bausubstanz. Die Dokumentation des Vorhabens erfolgt entsprechend entweder im Objektossier oder in einem neu eröffneten Projektossier (aus dem ebenfalls Unterlagen in das Objektossier kopiert werden können).

Das geschlossene Projektossier hat eine Aufbewahrungsfrist von null Jahren. Das Staatsarchiv übernimmt es bereits bei Inbetriebnahme des Gebäudes, beziehungsweise bei Abschluss des Umbau- oder Renovationsprojekts. Wie bisher liefert das Hochbauamt auch alle geschlossenen Objektossiers ab.

² Zumindest zu denjenigen, die das kantonale Eigentum in den letzten Jahrzehnten verließen.

³ Dies wäre auch bei einer analogen Dossierbewirtschaftung möglich gewesen, indem Plankopien zu „neuen Originalen“ hätten erklärt werden können. Allerdings wurden bei der Vervielfältigung von analogen Plänen über Jahrzehnte Verfahren angewendet, die für die Langzeitarchivierung ungeeignete Ergebnisse lieferten.

Neu werden Gebäudedossiers (Projekt- und Objektdossiers) nach einem einheitlichen Dokumentationsplan angelegt, bei dem es sich um eine vereinfachte Version der Empfehlung der Schweizerischen Koordinationskonferenz der Bau- und Liegenschaftsorgane der öffentlichen Bauherren KBOB handelt. Im Dokumentationsplan wird auch festgehalten, welche Unterlagen in welches Dossier – ins Projekt- und/oder Objektdossier – gehören.

Das Hochbauamt benutzt seinen Dokumentationsplan nicht nur „im Haus“, sondern hat ihn auch zur Grundlage des mannigfaltigen Austauschs mit den Dienstleistern der Baubranche sowie Behörden gemacht.

Wird ein Bauprojekt gestartet, eröffnet das Hochbauamt in seiner Datenbank „Planon“ ein nach dem Dokumentationsplan strukturiertes neues Projektdossier. Das Amt selber sowie die beauftragten Unternehmen speichern die betreffenden Dokumente gemäß dieser Struktur zentral ab. Die Speicherung auf dem Server erfolgt über den Webfileshare „Projektraum“. Dieses Interface erlaubt den Export von Gebäudedossiers und wird auch für die Ablieferung an das Staatsarchiv verwendet. Auch die Objektdossiers werden in der Planon-Datenbank gehalten und bis hin zur Archivierung über den „Projektraum“ verwaltet.

3 Vorteile für die Archivierung

Die Digitalisierung der Gebäudedossiers macht die Arbeit des Hochbauamts und der Unternehmen effizienter und transparenter. Auch für die Archivierung hat sie große Vorteile:

- Es werden sowohl das Gebäudedossier bei Inbetriebnahme als auch dasjenige bei Veräußerung oder Rückbau eines Gebäudes und damit Anfangs- und Endpunkt des betreffenden „Gebäude-Geschäftsfalls“ überliefert;
- Umbauten und Renovationen mit Projektcharakter und damit die wichtigsten baulichen Veränderungen werden ebenfalls überliefert;
- Es ist klar unterscheidbar, ob es sich bei einem Dossier um Bau- oder um Betriebsakten handelt;
- Vom Architekturbüro über das Hochbauamt bis ins Staatsarchiv ist nach einheitlichem Muster geregelt, welche Unterlagen wo abgelegt werden müssen;
- Das Hochbauamt muss die Gebäudedossiers für die Archivierung nicht extra ausdünnen. Mindestens die Triage der Unterlagen beim Übergang vom Projekt- zum Objektdossier macht das Hochbauamt aus eigenem Interesse sowieso.

4 Die Zukunft mit BIM und die Frage der Archivierung

Neben dem Umstieg auf das digitale Geschäftsdossier treibt das Hochbauamt die Digitalisierung auch in weiteren Bereichen voran. So sollen zukünftig alle kantonalen Hochbauprojekte mittels BIM (Building Information Modeling bzw. Bauwerksdatenmodellierung) in einem 3D-Gebäudemodell geplant, gebaut und bewirtschaftet werden.

Was kann man sich unter dieser Arbeitsmethode für die vernetzte Planung, den Bau und die Bewirtschaftung von Gebäuden konkret vorstellen? Bei der BIM-Methode werden die Bauwerksdaten (z. B. Daten zum Material, zur Position usw. von einzelnen Gebäudeteilen wie Wänden, Heizungen oder Lüftungen, aber auch Daten zum Wartungszyklus usw.) von sämtlichen Projektbeteiligten direkt in einem dreidimensionalen Gebäudemodell erfasst.⁴ Das virtuelle Gebäudemodell wird somit als strukturgebendes Element für alle Informationen rund um ein Bauwerk genutzt.

Die Innovation von BIM ist dabei nicht, dass mit 3D-Modellen gearbeitet wird – dreidimensional gezeichnete Modelle zur Visualisierung von Gebäuden werden von Architekten und Architektinnen schon seit längerer Zeit genutzt – sondern, dass diese Modelle für Koordinationsaufgaben und die Kommunikation innerhalb des Planungsteams genutzt werden. Die gemeinsame Arbeit an einem Modell bietet dabei zahlreiche Vorteile. So sind z. B. Änderungen (z. B. Einbau einer zusätzlichen Säule) und deren Auswirkungen (z. B. der fehlende Platz für das Öffnen des angrenzenden Fensters) direkt für alle Beteiligten ersichtlich.

Durch den Umstand, dass das Hochbauamt seine Gebäude zukünftig mittels BIM und ihren technischen Hilfsmitteln plant, baut und bewirtschaftet, stellt sich aus archivischer Sicht die Frage, ob die 3D-Gebäudemodelle archivwürdig und archivierbar sind.

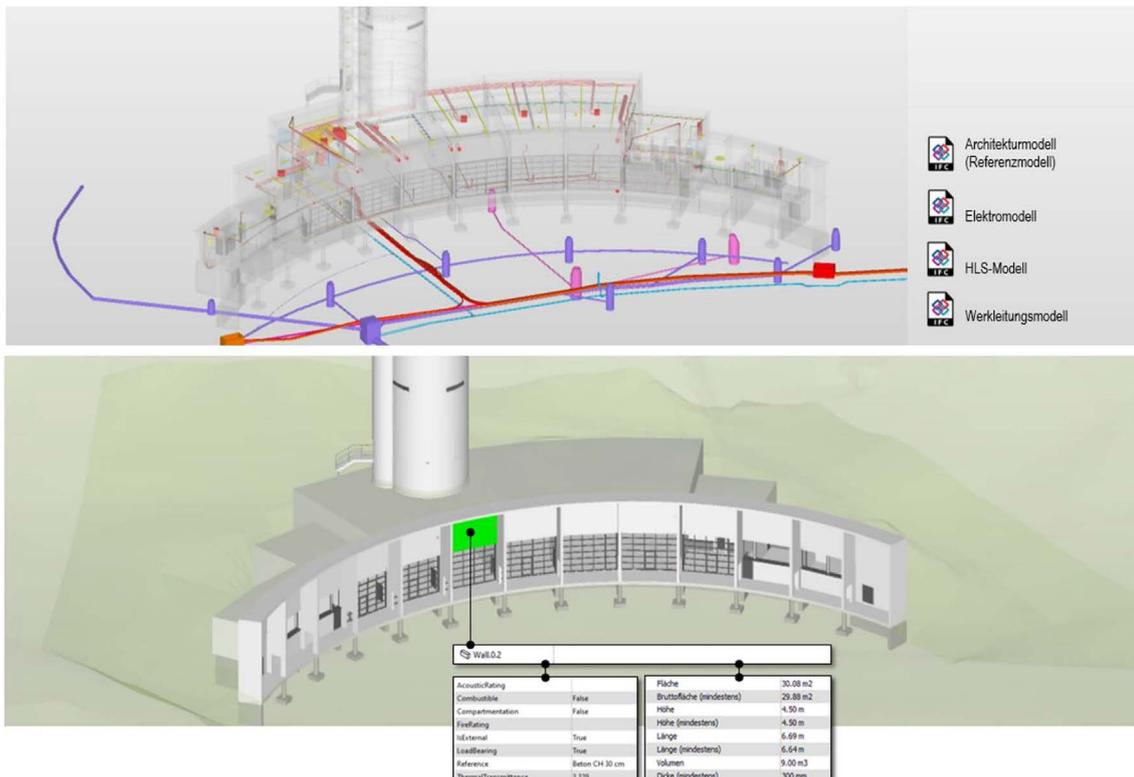
Bevor allerdings näher auf diese Fragen eingegangen wird, soll zunächst kurz anhand eines Beispiels gezeigt werden, wie die Entstehung und Nutzung eines 3D-Gebäudemodells beim Hochbauamt in der Praxis aussieht.

4 Die Aktualisierung der Daten erfolgt dabei allerdings nicht in Echtzeit, sondern erst nach der Kontrolle durch den Projektleiter bzw. die Projektleiterin (meist der Architekt oder die Architektin).

5 Das BIM 3D-Gebäudemodell in der Praxis des Hochbauamts

Die Grundlage des 3D-Gebäudemodells stellen verschiedene Teilmodelle im IFC-Format⁵ dar, die von den einzelnen Projektbeteiligten erarbeitet und vom Projektleiter / von der Projektleiterin zu einem Gesamtmodell zusammengeführt werden, wie es im oberen Teil in der untenstehenden Abbildung dargestellt wird.

Das 3D-Gebäudemodell besteht dabei einerseits aus geometrischen Daten und andererseits aus Elementen (wie z. B. Wänden, Fenstern oder Heizungsrohren), die ihrerseits Träger von strukturierten Daten sind. So ist im unteren Teil der untenstehenden Abbildung erkennbar, dass jedes einzelne Element Informationen zu Länge, Breite, Höhe, Gesamtfläche usw. enthält.

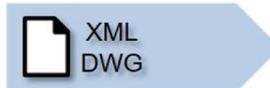
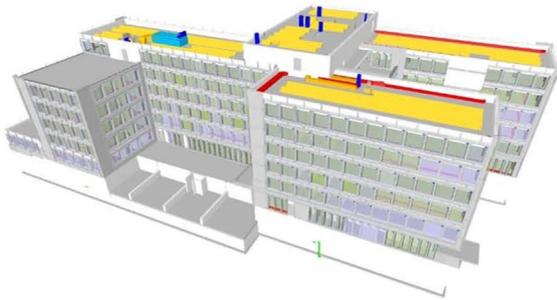


6 Das Verhältnis zwischen BIM 3D-Gebäudemodell und Projekt-/bzw. Gebäudedossier

Das BIM 3D-Gebäudemodell ist zwar Träger zahlreicher verschiedenartiger Informationen, ersetzt beim Hochbauamt jedoch nicht die Dokumentenablage im Projekt- bzw. Objektdossier in der eingangs bereits erwähnten Datenbank Planon. Vielmehr werden aus dem finalen BIM 3D-Gebäudemodell die relevanten Informationen wie Pläne usw. exportiert und ins Planon-Dossier importiert. So werden einerseits über XML-Exporte Metadaten – wie z. B. Angaben zur Größe eines Raumes – aus dem Gebäudemodell exportiert und in Planon importiert. Andererseits werden konventionelle Pläne aus dem Gebäudemodell generiert und in Planon abgelegt.

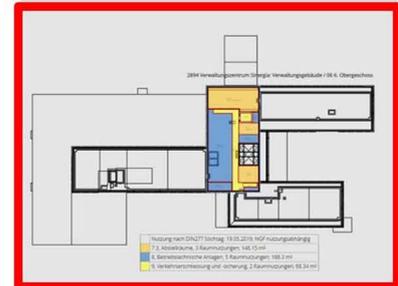
⁵ Industry Foundation Classes, Austauschformat für technische Gebäudeinformationen.

BIM - Koordinationsmodell



Datenbank-Planon

Objekt	
Objekt	2994_Verwaltungszentrum Sinergia_Verwaltungsgebäude
Stockwerk	06, 6. Obergeschoss
Raum	
Raumnr.	06.806
Raumbezeichnung	Haustechnik
Raumanschrift	Haustechnik
Reinigungsinformationen	
NGF Raumgröße	14.338 m ²
NGF Raumnutzung	14.338 m ²
Bodenfläche Raum...	14.338 m ²
Material Decke	Beton Typ 2
Material Wand	Beton Typ 2, KS, gestrichen
Material Boden	Monobeton
Raumnutzung	
Dienststelle	02.02, Hochbauamt
Nutzung nach DIN277	8, Betriebstechnische Anlagen
BIM GUID	d81c72cb-d80e-4181-8845-9a8157caae27



Aus der Planon heraus bestehen aber weiterhin Verknüpfungen zu den einzelnen Objekten im Gebäudemodell.

7 Das BIM 3D-Gebäudemodell und die Fragen für die Archivierung

Diese Praxis wirft in Bezug auf die Archivierung die Fragen auf, ob das BIM 3D-Gebäudemodell als archivwürdig einzustufen ist und wenn ja, ob es sich bei IFC um ein archivtaugliches Format handelt.

Für eine Bewertung des BIM-3D-Gebäudemodells als archivwürdig sprechen aus der Sicht der Autoren die folgenden Punkte:

- Das Gebäudemodell integriert sehr viele verschiedene und Disziplinen-übergreifende Informationen zu einem Bauwerk in einem zusammenhängenden Datenmodell.
- Das Gebäudemodell ist eher als „Original“ zu bewerten, während die exportierten Pläne in Planon eine Repräsentation davon sind.
- Die Archivierung des Gebäudemodells dokumentiert einen technischen Fortschritt und eine neue Arbeitsweise.
- Die Gebäudemodelle haben mit einer Größe von ca. 500 MB – 1 TB⁶ einen vertretbaren Datenumfang.

Gegen eine Bewertung des BIM-3D-Gebäudemodells als archivwürdig sprechen hingegen die folgenden beiden Punkte:

- Das Gebäudemodell dient primär als Planungs- und Verwaltungstool.
- Sämtliche im Gebäudemodell enthaltenen Informationen werden auch aus der Datenbank Planon abgeliefert.
- In Bezug auf die Frage, ob es sich beim IFC um ein archivtaugliches Format handelt, sehen die Autoren hauptsächlich zwei Punkte, die dafürsprechen:
 - Bei IFC handelt es sich um einen quelloffenen Standard.
 - IFC-Dateien sind menschenlesbar.

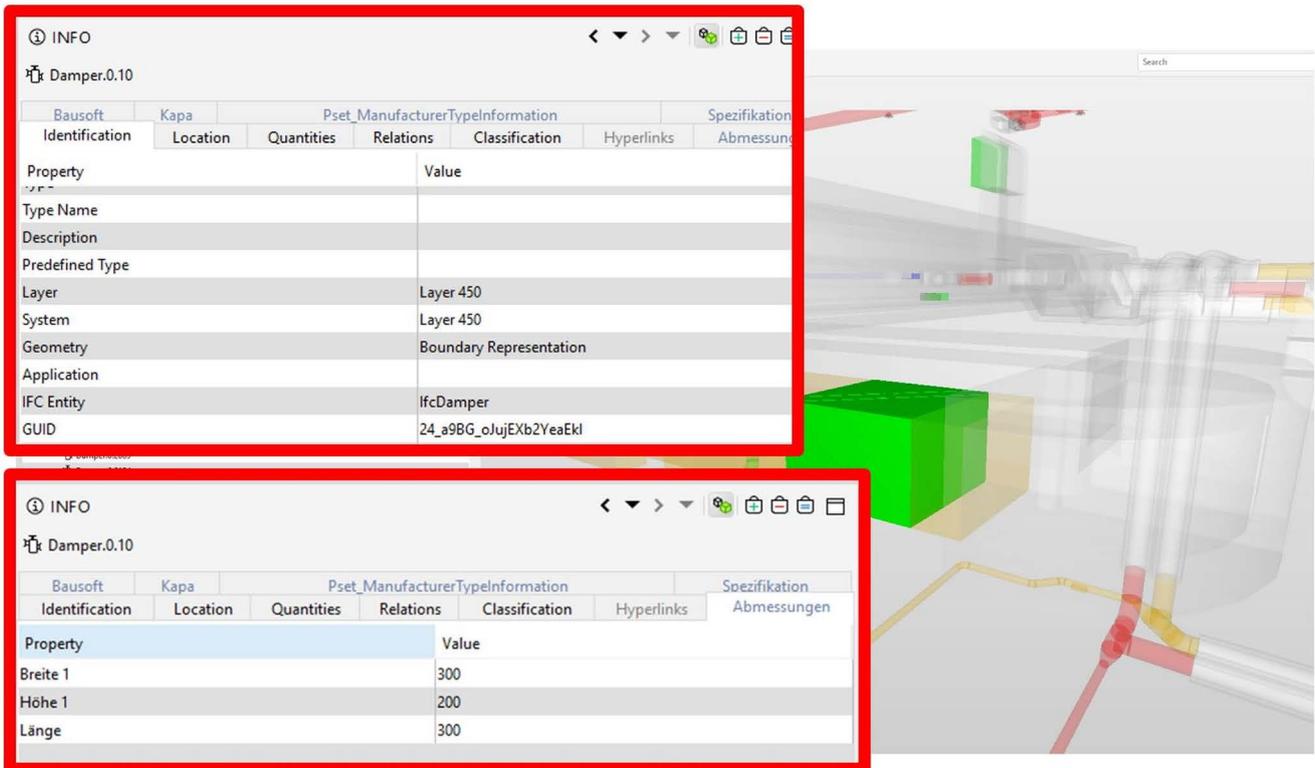
Dieser zweite Punkt soll abschließend an einem Beispiel ausgeführt werden.

⁶ Die Größe der Gebäudemodelle ist dabei weniger von der Größe des Gebäudes als vielmehr vom Detaillierungsgrad der Planung abhängig.

8 Menschenlesbarkeit des IFC-Formats: Ein Beispiel

Als Beispiel dient eine sogenannte Brandschutzklappe. Diese ist ein sicherheitstechnisches Bauteil, das aus Gründen des Brandschutzes zwischen Brandabschnitten in Wände und Decken eingebaut wird. Im Normalbetrieb ist die Brandschutzklappe geöffnet, um den Luftdurchlass zu ermöglichen. Sie schließt im Brandfall automatisch und verhindert so die Ausbreitung von Feuer und Rauch.

Brandschutzklappen sind Teil des Lüftungsmodells. Wie die untenstehende Abbildung zeigt, lässt sich jede Brandschutzklappe innerhalb dieses 3D-Modells eindeutig lokalisieren. Gleichzeitig ist jede Brandschutzklappe Trägerin von bestimmten Metadaten.



Sowohl die Metadaten der einzelnen Brandschutzklappe wie die eindeutige GUID (24_a9BG_oJujEXb2YeaEkl), die Bezeichnung (Brandschutzklappe) und die Längenmaße (Breite: 300) als auch ihre Lage innerhalb des Lüftungsmodells (0.20302935599037, -0.009693937276918, 0.0615000000000000) sind im IFC des Lüftungsmodells ganz einfach zu finden.

```
#334581=IFCDAMPER('24_a9BG_oJujEXb2YeaEkl',$,',',$,#334580,#334575,$,$)
#334599= IFCPROPERTYSINGLEVALUE
('Komponententyp'.$.IFCTEXT('Brandschutzklappe'),$)
#334585= IFCPROPERTYSINGLEVALUE('Manufacturer'.$.IFCTEXT('Schmidlin
AG'),$)
#334589= IFCPROPERTYSINGLEVALUE
('Breite1'.$,IFCREAL(300.000000000000000),#334588)
#334635= IFCCARTESIANPOINT ((0.203029355399037, -0.009693937276918,
0.0615000000000000))
```

9 Fazit

In vorliegendem Artikel wurde gezeigt, dass der gordische Knoten, den die Archivierung der analogen Gebäudedossiers des Bündner Hochbauamtes bis 2019 dargestellt hatte, dank der Digitalisierung durchschlagen werden konnte. Das analoge Gebäudedossier wurde während des gesamten Lebenszyklus' der betreffenden Hochbauten als offenes, laufend aktualisiertes und bereinigtes Geschäft geführt und dokumentierte deshalb zum Zeitpunkt seiner Archivierung lediglich den Zustand und die Nutzung des Gebäudes beim Ausscheiden aus dem kantonalen Besitz.

Seit 2020 wird pro Lebensphase eines Gebäudes ein digitales Dossier eröffnet, geführt und geschlossen. Dieses Vorgehen bietet für die Archivierung zahlreiche Vorteile. Der wichtigste davon ist, dass nun die verschiedenen Zeitstände aus dem Lebenszyklus eines kantonalen Gebäudes (Inbetriebnahme, Abbruch/Verkauf, größere Umbauten, Renovationen, Umnutzungen usw.) archiviert werden können.

Die Digitalisierung im öffentlichen Hochbau ist mit dem Umstieg auf das digitale Gebäudedossier noch keineswegs abgeschlossen. Eine wesentliche Neuerung ist der Umstieg auf BIM (Building Information Modeling bzw. Bauwerksdatenmodellierung). So werden künftig alle kantonalen Hochbauprojekte mittels BIM in einem 3D-Gebäudemodell geplant, gebaut und bewirtschaftet. Die Autoren haben in diesem Zusammenhang zur Diskussion gestellt, inwiefern diese BIM-3D-Gebäudemodelle als *archivwürdig* gelten können, ohne diesbezüglich eine abschließende Antwort liefern zu wollen. Unabhängig von dieser Frage erachten wir die Modelle als *archivierbar*. Dies, da die Modelle im quelloffenen IFC-Standard geschrieben sind und die IFC-Dateien menschenlesbar sind.

Gemeinsamer automatisierter Ingest archivwürdiger Fachverfahrensdaten im Kommunalen Digitalen Langzeitarchiv Baden-Württemberg

Miriam Eberlein, Stadtarchiv Heilbronn

Im Kommunalen Digitalen Langzeitarchiv Baden-Württemberg (künftig: KDL BW) nutzen derzeit (Stand Juli 2022) 95 Kommunal- und Kreisarchive eine gemeinsame DIMAG-Installation bei der Komm.ONE (AÖR). Zugleich sind die baden-württembergischen Kommunen Kunden bei der Komm.ONE für verschiedene Fachverfahrensanwendungen. Beim Personenstandswesen ist dies flächendeckend der Fall, beim Einwohnermeldewesen nahezu flächendeckend. Bei anderen Fachverfahren, beispielsweise beim Gewereregister, betrifft dies eine mittlere dreistellige Anzahl von Kommunen, während in den übrigen Kommunen Angebote anderer Dienstleister im Einsatz sind.

Für das Gewereregisterverfahren der Komm.ONE ging 2018 die Exportschnittstelle für zu archivierende Registerdaten produktiv, die zuvor in der baden-württembergischen AG Archivexporte entwickelt worden war.¹ Damit standen diejenigen Mitglieder des KDL BW, deren Kommunen dieses Fachverfahren nutzen, erstmals vor der Praxisfrage, wie sie die exportierten Gewereregisterdaten in ihr DIMAG übernehmen sollten. Da sich die Mitglieder des KDL BW in einem Anwenderkreis regelmäßig zu Fragen rund um DIMAG austauschen, kam der Vorschlag auf, in diesem Kreis eine gemeinsame Lösung für den Ingest der Gewereregisterdaten ins DIMAG zu entwickeln.

Der vorliegende Beitrag schildert die Voraussetzungen für diese gemeinsame Ingest-Lösung und skizziert, welche Vorteile, aber auch welche Einschränkungen sich daraus für die teilnehmenden Archive ergeben haben.

1 Ausgangslage und Ziel der gemeinsamen Ingest-Lösung

Mit der Produktivsetzung der Exportschnittstelle 2018 wird aus dem Gewereregisterverfahren regelmäßig einmal pro Jahr ausgesondert. Da Komm.ONE das Gewereregister, ebenso wie zahlreiche weitere Fachverfahren, als zentral gehostete Lösung anbietet, erfolgt dieser Export aus technischen Gründen für alle Kunden dieses Verfahrens zum gleichen Zeitpunkt. Die digitalen Aussonderungsportionen werden vom Betreiber Komm.ONE auf einer „Aussonderungsplattform“ zur Verfügung gestellt, in einer Art digitalem Postfach, sodass jedes Archiv jeweils nur auf die Daten derjenigen Kommune(n) Zugriff hat, für die es zuständig ist.

Um die exportierten Daten von der Aussonderungsplattform ins DIMAG zu übernehmen, muss entschieden werden, in welcher Struktur die Aussonderungsportion im DIMAG abgelegt werden soll. Außerdem können über ein Mapping mitgelieferte Metadaten direkt in DIMAG-Datenfelder eingespielt werden, sodass sie für die Recherche zur Verfügung stehen. Dieser Ingest-Prozess ist üblicherweise Sache eines jeden Archivs und wird von den DIMAG-Anwendern individuell gesteuert.

Die Intention einer gemeinsamen Ingest-Lösung war nun, diesen Ingest-Prozess möglichst effizient zu gestalten: Für die Aussonderungsportionen aus dem Gewereregister sollte ein Vorschlag für eine Ablage-Struktur im DIMAG sowie für das Metadaten-Mapping erarbeitet werden, der für alle DIMAG-Anwender nachnutzbar und für die regelmäßig jährlich zu erwartenden Ablieferungen wiederverwendbar war. Idealerweise sollte der Vorschlag auch auf Datenablieferungen aus Gewereregistern anderer Anbieter anwendbar sein, sofern sie bestimmte Voraussetzungen erfüllten. Zugleich sollte kein DIMAG-Anwender gezwungen sein, dem gemeinsamen Vorschlag zu folgen, sondern stets in der Lage bleiben, für sich selbst eine andere Ingest-Lösung zu realisieren.

In diese ersten Überlegungen war auch der für DIMAG zuständige IT-Fachmann bei der Komm.ONE eingebunden. Er stellte in Aussicht, bei hinreichend klar definierten Vorgaben für Ablagestruktur und Metadaten-Mapping eine Import-Routine zu entwickeln, die den jährlichen Ingest-Prozess automatisiert durchführen würde. Damit würde sich der manuelle Aufwand der DIMAG-Anwender auf die Sichtprüfung der Aussonderungsportionen und die jährliche Überprüfung des Ingest-Ergebnisses reduzieren.

1 Ausführlich zur AG Archivexporte (ehemals AG Archivschnittstellen) siehe: Katharina Ernst, Schnittstellen im Verbund. Zusammenarbeit der baden-württembergischen Stadt- und Kreisarchive mit dem DV-Verbund Baden-Württemberg, in: Kai Naumann, Peter Müller (Hrsg.), Das neue Handwerk: digitales Arbeiten in kleinen und mittleren Archiven. Vorträge des 72. Südwestdeutschen Archivtags am 22. und 23. Juni 2012 in Bad Bergzabern, Stuttgart 2013, S. 39–50. Zur Archivexport-Schnittstelle siehe den Vortrag von Katharina Ernst auf der 22. Tagung des Arbeitskreises zur Archivierung von Unterlagen aus Digitalen Systemen, Marburg 2018, online unter https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2018/praxis/20-03_Ingest__Ernst.pdf (abgerufen am 14.07.2023).

2 Die Voraussetzungen für einen automatisierten Ingest

Um die klaren Vorgaben für die Ablagestruktur und das Metadaten-Mapping zu treffen, musste der DIMAG-Anwenderkreis im Detail die folgenden beiden Fragekomplexe beantworten:

1. In welcher Struktur sollen die in der Aussonderungsportion (SIP) enthaltenen Dateien im DIMAG abgelegt werden?
DIMAG kennt vier Objektarten: Struktur-Objekt, Informationsobjekt (IO), Repräsentationsobjekt und File-Objekt. Sollte jede jährliche Ablieferung im DIMAG unter einem eigenen Struktur-Objekt abgelegt werden, oder genügte es, die SIPs unter einem Struktur-Objekt „Gewerberegister“ jeweils jährlich als neues Informationsobjekt abzulegen?
2. Welche der im SIP mitgelieferten Metadaten sollen in welche DIMAG-Metadatenfelder übernommen werden?

Auf der Grundlage eines vom Stadtarchiv Heilbronn erstellten Entwurfs wurden das Für und Wider verschiedener Varianten diskutiert und schließlich in einem Fachkonzept folgende Entscheidungen festgehalten:

Zu 1.) Struktur:

- Das SIP wird nicht aufgespalten, sondern zu genau einem AIP.
- Das AIP entspricht genau einem DIMAG-Informationsobjekt.
- Alle jährlichen Ablieferungen werden unter einem gemeinsamen, übergeordneten Struktur-Objekt „Gewerberegister“ abgelegt.
- Im DIMAG-Informationsobjekt wird genau eine Repräsentation angelegt, die sämtliche Dateien des SIP enthält.

Damit die jährlichen Exporte (SIPs) in jedem Archiv dem übergeordneten Strukturobjekt automatisiert zugeordnet werden können, muss jedes DIMAG-Anwenderarchiv dieses Strukturobjekt einmalig manuell anlegen und an Komm. ONE melden.

The screenshot shows the 'Strukturnavigation' interface. On the left, a tree view shows the hierarchy: D: Digitales Archiv > A: Archivalien > Mandant Heilbronn > StadtA HN: Stadtarchiv Heilbronn > B: Neuere Bestände > 028: Amt für öffentliche Ordnung > GEW: Elektronisches Gewerberegister. On the right, there is a search box 'Gehe zu AID:' with a 'go' button. Below the navigation is a table with columns 'Signatur/AID', 'Titel', and 'Details'. The table lists five entries for 'Gewerberegister, Ablieferung' with dates and times. A red box highlights a note: 'Anwender legt 1malig ein übergeordnetes Strukturobjekt für alle Ablieferungen an und meldet die AID an Komm.ONE.'

Signatur/AID	Titel	Details
StadtA HN B 028 GEW 1	Gewerberegister, Ablieferung 15.01.2018 10:00:10	🔑
StadtA HN B 028 GEW 2	Gewerberegister, Ablieferung 26.01.2018 10:29:10	🔑
StadtA HN B 028 GEW 3	Gewerberegister, Ablieferung 08.02.2018 14:34:38	🔑
StadtA HN B 028 GEW 4	Gewerberegister, Ablieferung 01.06.2019 00:28:32	🔑
StadtA HN B 028 GEW 5	Gewerberegister, Ablieferung 12.12.2020 09:41:35	🔑

Abb. 1: Definition der einheitlichen Struktur für den automatisierten Import von Aussonderungen aus dem Gewerberegister.
Vorlage: Stadtarchiv Heilbronn

Zu 2.) Metadaten-Mapping:

Nachdem mit der Definition der Struktur geklärt worden war, dass das DIMAG-Struktur-Objekt manuell, die anderen drei Objektarten automatisiert angelegt werden sollten, galt es, im zweiten Schritt für jede dieser drei Objektarten festzulegen, welche Metadatenfelder mit welchen Informationen vorbelegt werden sollten. Wie jede Datenbank, so kennt auch DIMAG einige (wenige) Pflichtfelder, die bei der Anlage eines Datensatzes nicht leer bleiben dürfen. Die Vorbelegungen sollten zudem die wesentlichen Informationen über die Aussonderungsportion abdecken, sodass auf manuelles Nacharbeiten grundsätzlich verzichtet werden kann. Dabei konnten die vorzubelegenden Informationen entweder statisch festgelegt sein („Konstante“), oder aber aus der in jedem SIP mitgelieferten Metadaten-Datei stammen, die wesentliche Informationen über die jeweilige Ablieferung enthält („welche Daten wurden wann, von wem und an wen geliefert“). Die Ergebnisse der Diskussion im DIMAG-Anwenderkreis wurden für jede DIMAG-Objektart in einer Tabelle festgehalten und gemeinsam beschlossen.

5.3 Informationsobjekt „Ablieferung“ (Ebenentyp = O)

Datenfeld in DIMAG	Vorbelegung	Inhalt	Datenquelle	Anmerkung
Archivialientyp*	JA	Fachverfahren	Konstante	In DIMAG: Auswahlliste
Signaturanteil	NEIN			Vorbelegung erscheint nicht sinnvoll, weil Archive unterschiedl. Signatursysteme haben. Jedes Archiv muss also nach dem Ingest diesen Signaturanteil selbst nachpflegen.
Titel*	JA	Gewerberegister, Ablieferung [DATUM]	Konstante; JJJJ-MM-TT aus metadaten.xml: <ErstellungsDatumUhrzeit>	Entscheidung des DIMAG-Anwenderkreises, hier vom ISAD(G)-Prinzip abzuweichen.
Beschreibung / Inhalt	JA	Datensätze der zwischen dem [TT.MM.JAHR] und dem [TT.MM.JAHR] abgemeldeten Gewerbe. Aussonderungsmetadaten.	Konstante; metadaten.xml <LoeschzeitraumVon> <LoeschzeitraumBis>	
Entstehungszeitraum*	JA	[TT.MM.JAHR – LoeschzeitraumVon] - [TT.MM.JAHR – LoeschzeitraumBis]	metadaten.xml: <LoeschzeitraumVon> - <LoeschzeitraumBis>	Anfangsdatum: Frühestmögliches Datum der Abmeldung eines Betriebs im Aussonderungszeitraum (= Inhalt von LoeschzeitraumVon). Enddatum: Letztmögliches Datum der Abmeldung eines Betriebs im Aussonderungszeitraum (= Inhalt von LoeschzeitraumBis). Ein anderes Anfangsdatum (z.B. älteste Datum der Anlage eines Gewerbedatensatzes) wäre nicht mit vertretbarem Aufwand zu ermitteln.
Abweichend dokumentierter Zeitraum	NEIN			
Benutzungshinweise	JA	Enthält personenbezogene Daten.	Konstante	
Rechte	NEIN			
Provenienz	JA	AGS: [AGS], [Gemeindenname], Gewerbebehörde	Konstante; metadaten.xml <LieferantGemeindeschlüssel>.	

Abb. 2: Beispiel aus der Tabelle des Metadaten-Mappings für das DIMAG-Informationsobjekt. Vorlage: Stadtarchiv Heilbronn

Nach der gemeinsamen Abstimmung des Fachkonzepts im DIMAG-Anwenderkreis erfolgte auf dieser Grundlage bei der Komm.ONE die Programmierung der automatisierten Import-Schnittstelle. Diese wurde anschließend für zwei Anwenderarchive erprobt und noch 2018 für den Produktiveinsatz freigegeben.

3 Ablauf der automatisierten Daten-Übernahmen ins DIMAG

Um am jährlichen automatisierten Ingest-Verfahren teilzunehmen, muss ein DIMAG-Anwenderarchiv die DIMAG-Anwenderbetreuung bei Komm.ONE einmalig beauftragen. Der Export von auszusondernden Daten aus dem Komm.ONE-Gewerberegister und ihr anschließender Import läuft daraufhin nach folgendem Schema ab:

Der Gewerberegister-Betreiber (hier: Komm.ONE) stößt den jährlichen Export an und erzeugt pro Gewerberegister-Kunde (Kommune) eine Aussonderungsportion. Diese wird auf der Aussonderungsplattform im digitalen „Postfach“ des zuständigen Kommunal- oder Kreisarchivs zur Einsicht zur Verfügung gestellt.

Die zuständigen Kommunal- oder Kreisarchive können innerhalb einer Frist von acht Wochen die Aussonderungsportionen prüfen. Diese Prüfungen sind optional und finden immer individuell und manuell statt. Ein Archiv kann beispielsweise die in der Aussonderungsportion enthaltenen XML-Dateien auf ihre Validität und Wohlgeformtheit hin checken. Die Vollständigkeit der Aussonderungsportion kann anhand einer elektronischen Ablieferungsliste, die auch die abliefernde Gewerbestelle erhält, kontrolliert werden. In Stichproben kann gemeinsam mit der Gewerbestelle die inhaltliche Vollständigkeit einzelner Datensätze überprüft werden, da die Daten erst nach Ende der achtwöchigen Frist aus dem Herkunftssystem gelöscht werden.

Soweit innerhalb der achtwöchigen Prüffrist die Aussonderungsportionen weder seitens der Archive noch seitens der Gewerbestellen beanstandet wurden, aktiviert der für DIMAG zuständige IT-Fachmann bei Komm.ONE die automatisierte Import-Schnittstelle. Die Aussonderungsportionen werden für jedes DIMAG-Anwenderarchiv in das zuvor angelegte und übermittelte Strukturobjekt verschoben. Dieser Verschiebung geht die automatisierte Überprüfung einiger technischer Parameter voraus: Hashwerte, Abwesenheit von 0-KB-Dateien, Vorhandensein aller zu erwartenden Dateien sowie ob alle im Mapping benannten Metadatenfelder ausgefüllt werden können oder ob gegebenenfalls in der Aussonderungsportion erwartete Metadaten fehlen.

Nimmt ein Kommunal- oder Kreisarchiv nicht am automatisierten Ingest-Verfahren teil, so kann es sich die Daten manuell von der Aussonderungsplattform herunterladen und anschließend für die Langzeitarchivierung bearbeiten.

4 Vorteile und Einschränkungen durch den automatisierten Import

Die Vorteile der automatisierten Import-Schnittstelle liegen auf der Hand: Im Prinzip können die jährlichen Ablieferungen der Gewerberegisterdaten nahezu ohne eigene Personalressourcen und DIMAG-Kenntnisse archiviert werden. Da sämtliche Prüfungen letztlich optional sind, genügt die einmalige Erstellung und Festlegung eines DIMAG-Strukturobjekts, um an dem automatisierten Import-Verfahren teilzunehmen. Damit ist es in Baden-Württemberg gelungen, die digitale Langzeitarchivierung der Gewerberegisterdaten in rund 400 Kommunen sicherzustellen, darunter zahlreiche kleine Gemeinden, die selbst nicht über geschultes Fachpersonal für diese Aufgabe verfügen.

Für diese Vorteile müssen die teilnehmenden Archive allerdings gewisse Einschränkungen in Kauf nehmen. Denn individuelle Abweichungen von der vereinbarten Vorlage für die Ablagestruktur und das Metadaten-Mapping sind nicht möglich. Dies ist für Archive eine neue Situation, waren sie doch in der Bewertung und Aufstellung ihrer Bestände bisher stets autonom. Durch die Festlegungen der automatisierten Import-Schnittstelle aber ist bereits entschieden, in welcher Struktur ein SIP im Archiv zum AIP transformiert wird.

Am Beispiel des Gewerberegisters zeigt sich, dass bereits diese Entscheidung weitere Weichenstellungen zwingend nach sich zieht: So wird das DIMAG-Informationsobjekt bei der Kopplung mit einem Archiv-Informationssystem (AFIS) stets als Verzeichnungseinheit definiert. Daraus ergibt sich also im Falle der Gewerberegisterdaten, dass die jährliche Ablieferung in der Tektonik eines Archivs zur Verzeichnungseinheit wird, während das übergeordnete Strukturobjekt dem Bestand oder zumindest der Serie innerhalb eines Bestandes entspricht. Auch die Verzeichnungsmetadaten, soweit sie das AFIS künftig automatisiert aus dem DIMAG übernehmen soll, werden bei der skizzierten automatisierten Ingest-Lösung archivübergreifend vereinheitlicht (wobei stets eine individuelle, manuelle Nachbearbeitung möglich bleibt).

Während die Vorgabe der Metadaten im Allgemeinen unproblematisch war, stieß die Vorgabe der Struktur bei einigen Anwendern in der Praxis auf Probleme: Denn zum einen setzt die vereinbarte Import-Struktur voraus, dass ein Archiv mit dem Konzept „offener“ Bestände oder Serien umgehen kann, also Bestände bzw. Serien, die nach und nach um weitere Verzeichnungseinheiten erweitert und somit niemals abgeschlossen werden. Zum anderen musste beispielsweise ein Kreisarchiv seine Tektonik anpassen, da es bislang die Archive der einzelnen von ihm betreuten Gemeinden als hierarchisch nicht weiter untergliederte Bestände geführt hatte. In dieser Tektonik, so stellte sich heraus, ließen sich die Gewerberegister-Ablieferungen nicht als „Unterbestand“ oder Serie unter das jeweilige Gemeindearchiv einordnen, weil diese dann entweder hierarchisch auf einer Stufe mit Verzeichnungseinheiten gestanden hätten oder aber auf gleicher Ebene mit den Gemeindearchiven selbst – in beiden Fällen keine zufriedenstellende Lösung.

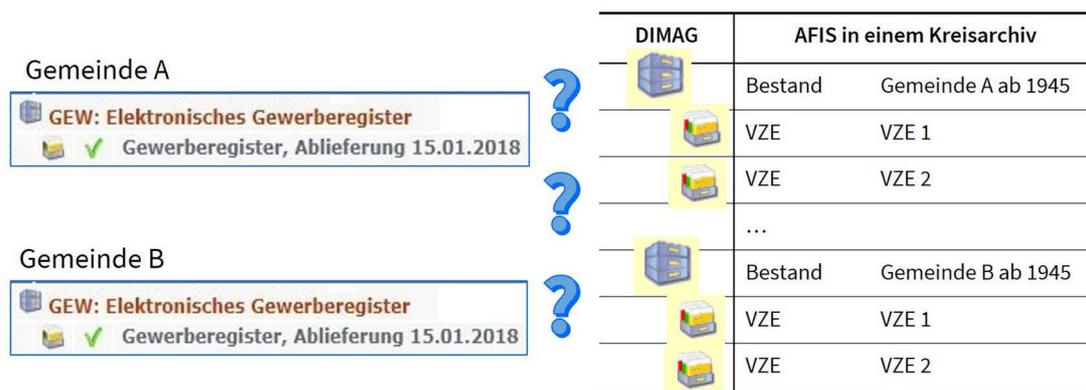


Abb. 3: Am Beispiel der Tektonik eines Kreisarchives zeigt sich, dass die Einordnung der mit dem automatisierten Ingest festgelegten Struktur nicht immer trivial ist. Vorlage: Stadtarchiv Heilbronn

5 Ausblick: Ausbau und Flexibilisierung der automatisierten Import-Schnittstelle

Seit 2018 werden nach dem geschilderten Verfahren jährlich die Daten aus dem Gewerberegister der Komm.ONE in das DIMAG der kommunalen Mandanten erfolgreich übernommen. Einige kommunale DIMAG-Anwender, deren Gemeinden für das Gewerberegister ein Fachverfahren anderer Hersteller einsetzen, haben inzwischen Interesse bekundet, ebenfalls die automatisierte Import-Schnittstelle zu nutzen. Dies ist grundsätzlich möglich, sofern folgende drei Voraussetzungen erfüllt sind:

- Die Dateinamen der im Export enthaltenen Dateien, ihre Anzahl und ihre Ablagestruktur müssen der im Import-Fachkonzept vorausgesetzten Benennungskonvention entsprechen (diese basiert auf dem Datenexport aus dem Komm.ONE-Gewerberegister).

- Es muss eine XML-Metadatendatei vorhanden sein, die Informationen über die Aussonderung in einer vordefinierten XML-Struktur enthält, da aus dieser Datei Informationen in die DIMAG-Metadatenfelder gemappt werden.
- Der Datenexport muss auf der Aussonderungsplattform der Komm.ONE abgelegt werden.

Diese Voraussetzungen sind im Fachkonzept für die Gewereregisterdaten-Import-Schnittstelle offengelegt, sodass sie Anbietern anderer Gewereregisterverfahren als Information zur Verfügung stehen.

Da bei der AG Archivexporte im nächsten Schritt die Fertigstellung der Export-Schnittstelle für die Einwohnermeldedaten ansteht, konzentrieren sich die derzeitigen Überlegungen für eine weitere automatisierte Import-Schnittstelle auf diese Datenart. Auf Basis der mit den Gewereregisterdaten gewonnenen Erfahrung wird dabei angestrebt, die Vorgaben zur Ablagestruktur seitens der Import-Schnittstelle zu flexibilisieren: Ziel ist es, dass die teilnehmenden Archive zwischen zwei Varianten für die Bildung des AIP, also des DIMAG-Informationsobjekts, auswählen können. Voraussetzung für diese Wahlmöglichkeit ist zunächst eine andere Zusammensetzung des SIPs: Dieses wird bei den Einwohnermeldedaten, anders als beim Gewereregister, nicht aus einer einzigen Primärdaten-XML-Datei bestehen, die sämtliche ausgesonderte Datensätze enthält. Stattdessen wird im SIP-Export aus dem Einwohnermelderegister jeder Primärdatensatz als einzelne XML-Datei abgelegt. Dies ermöglicht für die AIP-Bildung verschiedene Varianten: 1.) sämtliche Primärdaten-XML-Dateien in einem AIP zusammengefasst oder 2.) pro Primärdaten-XML-Datei ein separates AIP oder 3.) eine gruppenweise Zusammenfassung von Primärdaten-XML-Dateien, z. B. nach Anfangsbuchstabe des Nachnamens. Die Vor- und Nachteile der jeweiligen Varianten sollen hier nicht diskutiert werden – je nach Anzahl der zu erwartenden Primärdatensätze pro Ablieferung, nach Archivgröße und auch nach Vorlieben für eine flachere oder tiefere Erschließung wird hier jedes DIMAG-Anwenderarchiv seine eigene Entscheidung treffen müssen. Wichtig aber ist, dass es diese Entscheidung überhaupt treffen kann – denn sie hat, wie oben gezeigt, grundlegende Auswirkung darauf, wie letztlich die Verzeichnungseinheiten im AFIS des jeweiligen Archivs zusammengesetzt und in der Tektonik strukturell verankert sind.

Neben der Nutzung der automatisierten Import-Schnittstelle wird es selbstverständlich weiterhin jedem Mitglied des Kommunalen Digitalen Langzeitarchivs Baden-Württemberg freigestellt bleiben, sich für einen manuellen Import nach ganz eigenen Überlegungen zu entscheiden. Allerdings dürfte mit der Zunahme der Fachverfahren, die regelmäßig archivwürdige Daten exportieren, das Interesse der kommunalen DIMAG-Anwender an der Nutzung des automatisierten Import-Angebots wohl eher zu- als abnehmen. Schon bei den Gewereregisterdaten haben sich alle betroffenen Mitglieder des KDL BW für den automatisierten Import entschieden – für einige war das gar erst der Anlass zum Einstieg in die Digitale Archivierung. Die ja zumeist knappen Personalressourcen können so in andere mit der Digitalen Langzeitarchivierung verbundene Aufgaben investiert werden, für die (noch) keine Automatisierung in Aussicht steht.

Digitale Verbundarchive – Zu Chancen und Herausforderungen von archivischen Kooperationsmodellen in der digitalen LZA

Thomas Krämer, LVR-Archivberatungs- und Fortbildungszentrum, Pulheim

2005 sprach der Mannheimer Archivar Christoph Popp in einem mittlerweile mehrfach zitierten Aufsatz von katastrophalen Ereignissen wie „Flut oder Dürre“, mit denen Archivar/innen die digitale Langzeitarchivierung des Öfteren assoziierten.¹ Dahinter liegt wahrscheinlich zum Teil die Befürchtung, entweder von dieser Mammutaufgabe gleich einem elektronischen Tsunami überrollt zu werden. Oder aber, das Archiv wird von den digitalen Entwicklungen aus unterschiedlichen Gründen abgeschnitten und die Einrichtung in der Folge mittel- oder langfristig abgehängt. Die Beschäftigung mit der Herausforderung, die die digitale Langzeitarchivierung insbesondere für kleinere und mittlere Einrichtungen mit sich bringt, ist also nicht neu und wurde in verschiedenen Facetten diskutiert.² Einen Lösungsansatz bieten Entwicklungs- und Nutzer-Verbünde wie DIMAG, das DA NRW oder das Elektronische Kommunalarchiv (eKA) in Sachsen.³ Sie gehen die digitale Langzeitarchivierung kooperativ an und verteilen die Aufgaben und Lasten auf viele Schultern. Während auf diese Weise oder mit stand-alone-Lösungen viele staatliche wie nichtstaatliche Archive bereits in der digitalen Langzeitarchivierung aktiv sind und zunehmend produktive Lösungen einsetzen, bleibt das Aufgabenfeld gerade für mittlere und kleinere Einrichtungen eine enorme finanzielle, personelle, organisatorische und auch fachliche Herausforderung.⁴ Selbst die Verbünde können allerdings nicht alle Probleme gerade mittlerer und kleinerer Einrichtungen angehen oder sogar lösen. Ein weiterer, komplementärer Ansatz kann in der Schaffung von Verbundarchiven im Sinne von archivischen Gemeinschaftseinrichtungen liegen. Da aber auch sie nicht voraussetzungslos sind, sollen im Folgenden verschiedene Herausforderungen näher betrachtet werden. Einige Aspekte werden nur angerissen und nicht abschließend behandelt. Der Fokus liegt dabei auf der Situation in Nordrhein-Westfalen, was nicht bedeutet, dass die Ergebnisse nur in diesem Bundesland gültig wären.

Als Verbundarchiv wird im Weiteren eine Einrichtung verstanden, in der sich mehrere Archive bzw. Träger zur gemeinsamen Erledigung von Fachaufgaben zusammenfinden, wobei Umfang und Art der Integration sowie der Aufgaben- bzw. Lastenverteilung stark differieren können. Eine solche Verbundeinrichtung ist auch archivspartenübergreifend denkbar, allerdings in der Praxis bislang unüblich. Die Idee ist im Grunde nicht neu. Es existieren z. T. bereits seit langen Jahren unterschiedliche Kooperationsformen, die allerdings bislang v. a. auf den analogen Bereich konzentriert waren bzw. sind.⁵ Betrachtet man die Merkmale Personal, Infrastruktur und Haushalt lassen sich abstrahierend drei Gruppen bilden, deren Integrationsgrad variiert:

1. Minimalintegration: Das Verbundarchiv nutzt gemeinsame Infrastruktur, z. B. dasselbe Gebäude oder Magazine, aber die Verbundpartner unterhalten eigenes Personal, das getrennte Haushalte bewirtschaftet. Die Kooperation erstreckt sich nur auf die äußere, bauliche Hülle.
2. Teilintegration: Das Verbundarchiv setzt gemeinsames Personal ein, aber jeder Verbundpartner unterhält noch eigene Infrastruktur und Haushalt. Die Integration vertieft sich, da das Personal Zugriff auf das Registratur- und Archivgut aller Kooperationspartner hat.
3. Vollintegration: Das Verbundarchiv setzt gemeinsames Personal sowie Infrastruktur ein und verfügt über einen einzigen Haushalt. Dies erfordert eine tiefgreifende und umfassende Regelung aller Aspekte der Archivierung.

1 Christoph Popp, Abzugsgraben oder Überschwemmung – Strategien eines Kommunalarchivs für die dauerhafte Archivierung digitaler Unterlagen, in: Archive in Thüringen. Sonderheft 2005: Bewertung und Bestandsergänzung, Weimar 2005, S. 93.

2 Beispielsweise wurde 2012 die Einrichtung einer bundesweit agierenden Beratungsstelle, die mit der Schweizer KOST vergleichbar gewesen wäre, erwogen. Siehe hierzu: Ulrich Schludi, Brauchen wir eine Koordinierungsstelle für die digitale Archivierung? Zusammenfassung der Diskussion [nestor-Workshop „Brauchen wir Koordinierungsstellen für die digitale Archivierung?“], in: Christian Keitel (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus Digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“. Stuttgart 2013, S. 311–316.

3 Siehe zu den Verbänden zusammenfassend und vergleichend: Elisabeth Klindworth, Kooperative Verbünde zur digitalen Archivierung in der Praxis. Masterarbeit FH Potsdam. Potsdam 2020, online unter: <https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/2474> (abgerufen am: 14.07.2023).

4 Für eine grundsätzliche Analyse möglicher Kooperationsmodelle siehe: KOST-Projekt 17-035: Szenarien und Möglichkeiten einer DLZA im Verbund, Redaktion Karola Brüggemann und Pascal Föhr, Bern 2018: https://kost-ceco.ch/cms/dl/77b99cd251a61c7fb1a3a530cfd565/17-035_Verbund_Abschlussbericht_v1.0.pdf (abgerufen am: 14.07.2023). Ulrich Fischer hat eine Typologie von Verbänden vorgelegt: Gemeinsame Lösungen für ein gemeinsames Problem. Verbundlösungen für die elektronische Langzeitarchivierung in Deutschland, in: Archivpflege in Westfalen und Lippe 80 (2014), S. 21f.

5 Zur Situation im Rheinland siehe: Johannes Stinner, Archivverbünde in der rheinischen Kommunalarchivlandschaft, in: Claudia Kauertz (Red.): Archivlandschaft Rheinland. 49. Rheinischer Archivtag, 18.–19. Juni 2015 in Pulheim-Brauweiler. Beiträge, Bonn 2016, S. 95–101. Zu Westfalen siehe: Gunnar Teske, Gemeinsam sind wir stark. Beispiele interkommunaler Zusammenarbeit im Archivwesen, in: Archivpflege in Westfalen und Lippe 89 (2018), S. 23–28. Für Thüringen liegt ein Mustervertrag der Archivberatungsstelle vor: <https://landesarchiv.thueringen.de/media/landesarchiv/2Aufgaben/Archivberatungsstelle/gemeinschaftseinrichtungen.pdf> (abgerufen am: 14.07.2023).

Jede diese Kooperationsformen könnte auch die digitale Archivierung angehen. Neu ist jedoch die zusätzliche Möglichkeit, dass Archive, die bei analogen Unterlagen und AV-Medien unabhängig bleiben, ausschließlich im Bereich der digitalen LZA kooperieren und ein rein digitales Verbundarchiv gründen. Diese Optionen wurden schon früh erkannt, da verschiedene Kooperationsmodelle bereits in das OAIS-Referenzmodell (Magenta Book) eingingen. Die Autoren beschäftigten sich mit der Interoperabilität von Archiven und skizzierten mehrere denkbare Varianten.⁶ Im Jahr 2008 legte eine Arbeitsgruppe der International Federation of Library Associations and Institutions einen Bericht vor, der sich ebenfalls intensiv mit Kooperationsformen befasste.⁷

Die gesetzliche Grundlage für Verbundarchive ist in den meisten Bundesländern bereits vorhanden. Sie beruht in NRW zum einen auf dem Gesetz über kommunale Gemeinschaftsarbeit (GkG NRW) bzw. dessen Äquivalenten in anderen Bundesländern.⁸ Zum anderen sehen die meisten Landes-Archivgesetze ausdrücklich die Schaffung von Gemeinschaftseinrichtungen vor.⁹

Einige wichtige Kostenfaktoren der digitalen LZA lassen sich bereits durch die Teilnahme an einer Verbundlösung reduzieren.¹⁰ So fällt – je nach Konstellation des Verbundes – weniger Aufwand z. B. in den Bereichen IT-Sicherheit, Systembetreuung und -wartung, Software und Lizenzen, Technology Watch und Weiterentwicklung an.¹¹ Auf weitere Aspekte wie Personal wirkt sich die Teilnahme an einem Verbund zwar ebenso positiv aus, allerdings in geringerem, z. T. unterschiedlichem Ausmaß. Beispielsweise bindet die Vorfeldarbeit nach wie vor große Personalkapazitäten. Auch die Einarbeitung in die komplexe Materie und der Produktivbetrieb einer technischen Lösung setzen voraus, dass fachliche Kapazitäten aufgebaut werden können. Vorteile entstehen eher durch die gemeinsame Entwicklung oder Nachnutzung von Konzepten und Schnittstellen, für deren Erarbeitung weniger Ressourcen eingesetzt werden müssen. Für kleinere Einrichtungen kann somit selbst die Teilnahme an einer Verbundlösung wie DIMAG oder DA NRW eine große Herausforderung darstellen.

Als Reaktion haben baden-württembergische Kommunen die Option, über die sog. „skalierbare technische Zusammenarbeit“ an DIMAG zu partizipieren. Damit wird Kreisarchiven die Möglichkeit eröffnet, die digitale LZA für kreisangehörige Kommunen zu übernehmen, die diese Aufgabe nicht selbst stemmen können oder wollen. Der jeweilige Leistungsumfang und die Kostenverteilung werden zwischen Kreis und Kommune geregelt. Für kleine Kommunen existieren Sonderregelungen, sofern Daten aus Fachverfahren automatisiert übernommen und bei einem teilnehmenden Archiv abgelegt werden können.¹² Dieses Vorgehen spiegelt insofern die Situation in Baden-Württemberg wider, als den Kreisarchiven die Archivpflege in ihrem Sprengel obliegt. Beratungsstellen für nichtstaatliche Archive wie in Brandenburg, Hessen und Nordrhein-Westfalen existieren mit Ausnahme der DIMAG-Supportstelle beim Landesarchiv nicht. Das Vorgehen bietet gerade kleineren Archiven einen niederschweligen, kooperativen Ansatz und gewährleistet die Flexibilität, regionale oder lokale Besonderheiten berücksichtigen zu können.

- 6 Consultative Committee for Space Data Systems (Hrsg.), Reference Model for an Open Archival Information System (OAIS). Recommended Practise CCSDS 650.0-M-2, Washington D.C. 2012, Kap. 6, online unter: https://cwe.ccsds.org/cesg/docs/Previous%20CCSDS%20Organization/Panel%202%20-%20Information%20Interchange%20Services/Historical%20Documents/CCSDS%20650.0-R-2_%20Reference%20Model%20for%20an%20Open%20Archival%20Information%20System.pdf (abgerufen am: 14.07.2023).
- 7 International Federation of Library Associations and Institutions (Hrsg.), Public Libraries, Archives and Museums: Trends in Collaboration and Cooperation. Den Haag 2008; online unter: <https://www.ifla.org/files/assets/hq/publications/professional-report/108.pdf> (abgerufen am: 14.07.2023).
- 8 In den §§ 1 und 23 GkG NRW sind die grundsätzlichen Kooperationsformen sowie die Möglichkeit einer Aufgabenübertragung geregelt. Siehe Gesetz über die kommunale Gemeinschaftsarbeit NRW, online unter: https://recht.nrw.de/lmi/owa/br_bes_text?bes_id=3978&aufgehoben=N&anw_nr=2 (abgerufen am: 14.07.2023). Ähnlich zum Beispiel das baden-württembergische Gesetz über die kommunale Zusammenarbeit (GKZ), in §§ 1 und 25; online unter: <http://www.landesrecht-bw.de/jportal/?quelle=jlink&query=KomZG+BW+Inhaltsverzeichnis&pml=bsbawueprod.pml&max=true> (abgerufen am: 14.07.2023).
- 9 Siehe: Gesetz über die Sicherung und Nutzung öffentlichen Archivguts im Lande Nordrhein-Westfalen, § 10, 2: https://recht.nrw.de/lmi/owa/br_text_anzeigen?v_id=1000000000000000338 (abgerufen am: 14.07.2023). Ähnlich z. B. in Baden-Württemberg, Brandenburg, Mecklenburg-Vorpommern, Niedersachsen, Rheinland-Pfalz, Sachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen.
- 10 Zu den Kostenfaktoren der dLZA liegt noch keine abschließende Übersicht vor. Modelle wie LIFE oder DP4Lib lassen sich nur teilweise auf das Archivwesen übertragen. Siehe: Christina Starkloff, Übertragbarkeit des Kostenmodells zur Langzeitarchivierung LIFE auf den archivischen Bereich. Transferarbeit Archivschule Marburg, Marburg 2013; online unter: https://www.landesarchiv-bw.de/sixcms/media.php/120/55273/Transferarbeit2013_Starkloff.pdf (abgerufen am: 14.07.2023); Michael Ucharim, DP4lib als Kostenmodell für die digitale Langzeitarchivierung im Archivwesen? Eine Fallstudie am Beispiel des „Digitalen Archivs des Landes Hessen“; online unter: https://www.landesarchiv-bw.de/sixcms/media.php/120/55275/Transferarbeit2013_Ucharim.pdf (abgerufen am: 14.07.2023). Das EU-Projekt 4C beschäftigte sich ebenfalls mit der Frage der Kosten in der digitalen LZA: <https://www.4cproject.eu> (abgerufen am: 14.07.2023).
- 11 Die Auflistung erhebt keinen Anspruch auf Vollständigkeit. Zu Kostenfaktoren siehe: Julia Krämer-Riedel, Tobias Schröter-Karin: Archivierung im Verbund. Kosten der digitalen Langzeitarchivierung am Beispiel von DiPS.kommunal, in: Informationswissenschaft. Theorie, Methode und Praxis 5, 1 (2018), S. 49.
- 12 Siehe hierzu: Christian Keitel, Miriam Eberlein, Manfred Waßner, „DIMAG“ wird kommunal. Ein digitales Langzeitarchiv für Städte und Gemeinden in Baden-Württemberg, in: Generaldirektion des Österreichischen Staatsarchivs (Hrsg.), Digitale Archivierung. Innovationen – Strategien – Netzwerke. Innsbruck, Wien 2016, S. 21–31.

In Nordrhein-Westfalen ging das DA NRW einen anderen Weg, was u. a. auf die spezifischen Strukturen zurückzuführen ist. Position, Ausstattung und Struktur der Kreisarchive in NRW variieren stark. Sie reicht von größeren Kreiszentralarchiven, die die Unterlagen einiger oder sogar aller kreisangehörigen Kommunen archivieren, bis hin zu Einrichtungen, die nur die Überlieferung des Kreises selbst übernehmen. Dementsprechend divergiert die personelle und finanzielle Ausstattung, was sich letztendlich auf die Zusammenarbeit mit umliegenden Archiven auswirkt.¹³ Das Bundesland weist eine dichte archivische Infrastruktur auf. Hinzu kommt, dass mit den Archivberatungsstellen der Landschaftsverbände Westfalen-Lippe und Rheinland zentrale Ansprechpartner zur Verfügung stehen. Diese Faktoren tragen dazu bei, dass Verbundarchive nicht ausschließlich als Anschluss eines kommunalen Archivs an ein größeres (Kreis-)Archiv gedacht werden sollten. Vielmehr käme ebenso ein Zusammenschluss mehrerer kleinerer Einrichtungen in Betracht, ohne dass einem Archiv eine klare Führungsrolle in allen Bereichen zukommen muss.

Vor diesem Hintergrund wurde im Digitalen Archiv NRW (DA NRW) die Option einer sog. „Mandantenzusammenlegung“ entwickelt, die sowohl für die DA NRW Software-Suite (DNS) als auch für DiPS.kommunal gilt. Sie sieht im Kern die Nutzung eines Mandanten durch mehrere Einrichtungen eines Verbundarchivs vor.¹⁴ Damit steht sie auf den ersten Blick im Widerspruch zur Mandantenfähigkeit eines digitalen Archivs, die bei Verbundlösungen eine wichtige Rolle spielt. Dieser basale Baustein der Software-Architektur setzt letztlich eine datenschutzrechtliche Kernforderung um und ist entscheidend für die Integrität und Vertraulichkeit von digitalem Archivgut.¹⁵ Alle betrachteten Verbundlösungen betonen dementsprechend die Mandantenfähigkeit als essenzielles Merkmal für eine vertrauenswürdige und rechtskonforme Auftragsdatenverarbeitung. Im Zusammenhang von Verbundarchiven führt es aber zu erheblichen Mehrkosten, wenn für jeden Verbundpartner ein eigener Mandant eingesetzt werden müsste. Übertragen auf die bislang praktizierten Kooperationsmodelle entspräche dies getrennten Magazinen für jedes teilnehmende Archiv. Gerade bei Kooperationen fällt damit einer der wichtigsten Vorteile, die Reduktion von Kosten, erheblich geringer aus. Verpflichten sich jedoch mehrere Einrichtungen im Vorfeld auf eine gemeinsame Nutzung eines Mandanten, akzeptieren damit die Konsequenzen und treten als ein Auftraggeber auf, bleibt die strikte Trennung von anderen Mandanten unberührt. Damit wäre die Mandantenfähigkeit der eingesetzten Lösung nicht gefährdet.

Sowohl das baden-württembergische Modell, als auch die Mandantenzusammenlegung in NRW lassen mehrere Kooperationsmodelle zu. Zieht man die Verbundarchive in Betracht, die bereits für analoge Unterlagen bestehen, bilden sich mehrere, zunächst idealtypische Konstellationen heraus:

- 13 Gregor Patt, *Gemeinsam für eine starke Infrastruktur – Kreisarchive und LVR-AFZ als Partner bei der Archivberatung*, in: *Eildienst 4* (2019), S. 252f.
- 14 Es besteht natürlich keine Verpflichtung aller an einem Archivverbund teilnehmenden Einrichtungen, einen gemeinsamen Mandanten zu nutzen. Es ist durchaus denkbar, dass gerade größere Verbundpartner einen eigenen Mandanten nutzen und sich mehrere kleinere Einrichtungen einen weiteren Mandanten teilen. Insgesamt würde der Archivverbund somit mehrere Mandanten einsetzen.
- 15 Siehe hierzu: Bundesamt für die Sicherheit in der Informationstechnik (Hrsg.), *IT-Grundschutz-Kompendium* (Stand Februar 2020), Köln 2020, S. 6: „Als mandantenfähig werden Anwendungen, IT-Systeme oder auch Dienstleistungen bezeichnet, bei denen die Prozesse, Informationen und Anwendungen eines Mandanten strikt von denen anderer Kunden getrennt sind, also keine Zugriffe oder Störungen von dem einen in den anderen Bereich möglich sind und somit auch deren Vertraulichkeit, Integrität oder Verfügbarkeit nicht beeinträchtigt werden kann.“

1. Mehrere Archive, die jeweils keine oder nur sehr geringe Stellenanteile in die Langzeitarchivierung investieren können, kommen überein, diese u. U. zusätzlichen Stellenanteile zu addieren. Die Kooperation bestünde nur aus der gemeinsamen Stelle, die für alle beteiligten Einrichtungen nach einem zuvor definierten Schlüssel tätig wird. Die Archive könnten jedoch eigene Mandanten betreiben. Der Vorteil liegt in der gebündelten Aufgabenwahrnehmung durch eine zentrale Stelle, während die sonstige Unabhängigkeit der Institutionen gewahrt bleibt. Durch die Schwerpunktbildung kann sich diese Person eingehender mit der Aufgabe beschäftigen, was sich positiv auf die Aufgabenerledigung auswirkt. Diese gemeinsame Stelle erhält jedoch Einblick in die digitalen Unterlagen aller Kooperationspartner.

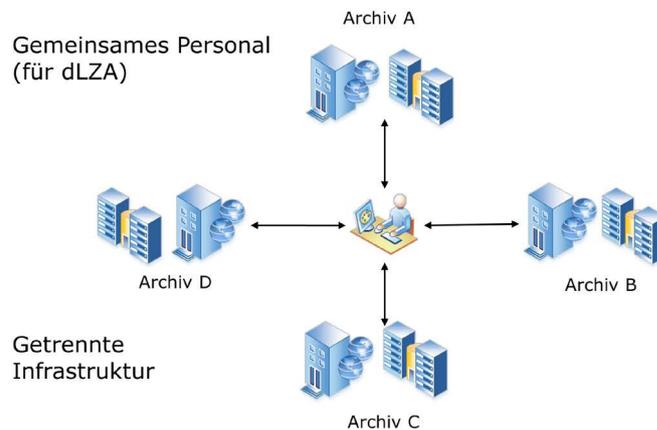


Abb. 1: Denkbare Option I

2. Ein Verbundarchiv könnte einen gemeinsamen Mandanten nutzen, jedoch bleiben Personal und Haushalte den einzelnen Verbundpartnern zugeordnet. Der Vorteil der Variante liegt v. a. in den verteilten Kosten für den Mandanten, während sonstige Kostenfaktoren nur wenig oder gar nicht berührt werden. Auch in diesem Szenario wird die Autonomie der einzelnen Archive weitgehend gewahrt. Alle Berechtigten der Verbundpartner können jedoch das digitale Archivgut des gesamten Mandanten einsehen.

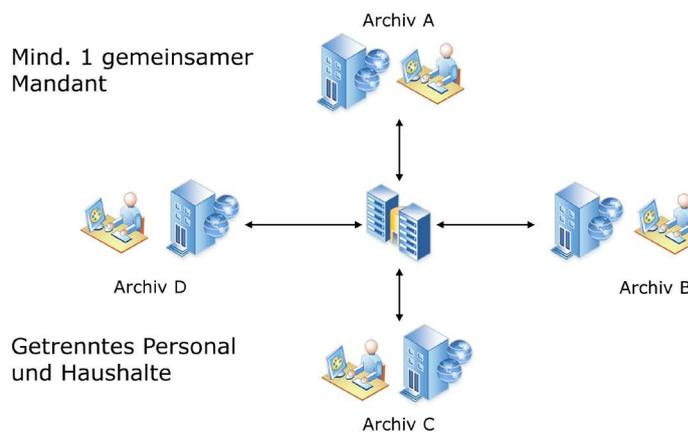


Abb. 2: Denkbare Optionen II

3. Die Varianten 1 und 2 könnten kombiniert werden: Zusätzlich zu einem gemeinsamen Mandanten könnte ein Verbundarchiv gemeinsames, gegebenenfalls zusätzliches Personal für die digitale Langzeitarchivierung und Behördenberatung einstellen. Die Kraft bzw. Kräfte würden mit einem Stellenanteil für die jeweiligen Verbundpartner tätig sein. So könnten nicht nur die Kosten gesenkt, sondern darüber hinaus auch eine einheitliche Beratung und ein Wissenstransfer sichergestellt werden.

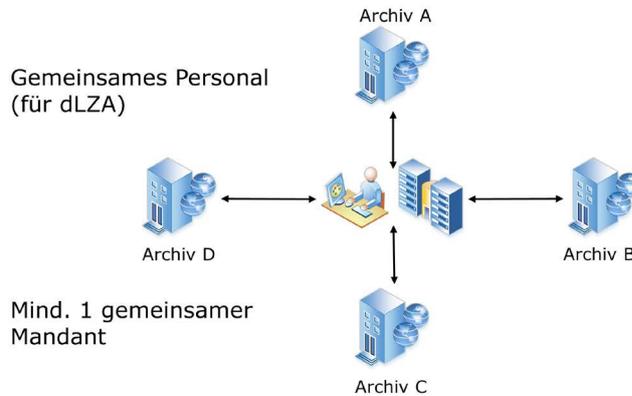


Abb. 3: Denkbare Optionen III

4. Die Verbundpartner kommen überein, Mandant, Haushalt und Personal für die digitale Langzeitarchivierung bei einem Partner oder einer neuen Einrichtung vollständig zu verankern. Der Partner betreibt einen Mandanten, hat das Personal in seinem Stellenplan und verfügt über die Haushaltsmittel. Damit geben die übrigen Verbundpartner einen guten Teil ihrer Autonomie auf und schaffen ein digitales Zentralarchiv, wohingegen die analoge Überlieferung in weiterhin in eigenständigen Institutionen verbleiben könnte.¹⁶

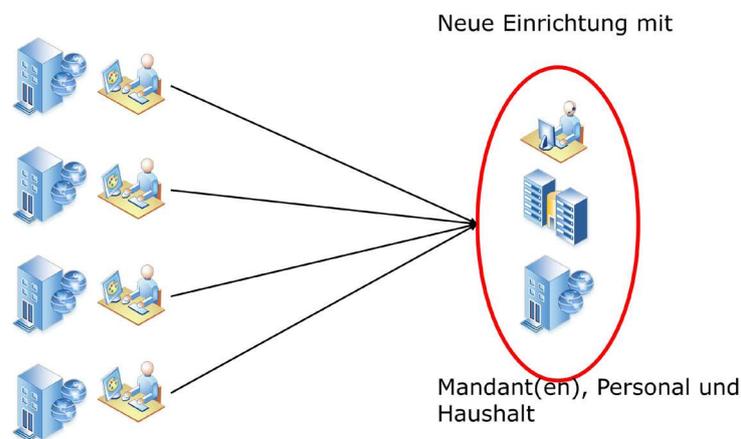


Abb. 4: Denkbare Optionen IV

Natürlich sind Mischformen dieser idealtypischen Varianten vorstellbar. Sollte ein Verbundmitglied einen eigenen Mandanten bevorzugen, könnten andere Partner immer noch einen gemeinsamen abrufen. Ähnliche Konstellationen wären sogar hinsichtlich des Personals denkbar, was aber die Komplexität deutlich erhöhen und Praktikabilität signifikant verringern würde.

Ungeachtet der Konstellation gibt es Aspekte, die auf jeden Fall in einer öffentlich-rechtlichen Vereinbarung verbindlich geregelt werden sollten, um spätere Komplikationen und Missverständnisse zu vermeiden. Sollte die Kooperation nicht schon auf anderen Feldern eingeübt sein, ist es u. U. für einen institutionalisierten Austausch zwischen den Trägern vorteilhaft, einen Beirat o. ä. zu bestimmen. Für Konfliktfälle, die auch in diesem Gremium nicht beigelegt werden können, ist es sinnvoll, weitere Vereinbarungen die Art und Weise der Klärung von Rechtsfragen zu treffen. Zu den Rechtsfragen gehört ebenso die Klarstellung, dass die Schaffung eines Verbundarchivs ungeachtet der Integrationstiefe eigentums- und urheberrechtliche Fragestellungen nicht berührt. Weitere rechtlich relevante Regelungen wie die Erarbeitung einer Satzung, einer Lesesaal-, Nutzungs- oder Gebührenordnung sowie deren Aktualisierung sind zu bedenken. Werden auf dieser

¹⁶ Inwiefern getrennte Teilarchive eines Trägers für bestimmte Materialitäten sinnvoll oder gewünscht sind, bleibt dahingestellt.

Basis Einnahmen generiert, ist nicht nur deren Verwendung zu bedenken, sondern auch die Frage, wer gegebenenfalls ausstehende Forderungen eintreibt. Wird Personal gemeinsam genutzt, macht es nicht nur Sinn, die Stellenanteile und die Eingruppierung einschließlich der damit verbundenen Kosten zu definieren. Auch personalrechtliche Gesichtspunkte sind zu beachten. Dazu gehören die fachliche wie disziplinarische Anbindung ebenso wie das Verfahren zur Stellenbesetzung und gegebenenfalls Beförderung. Zu diesem Komplex zählt auch die Zusammenarbeit bzw. die Rolle der übrigen Mitarbeitenden in den beteiligten Archiven, um Verantwortlichkeiten klar festlegen zu können. Die Verteilung der Kosten, z. B. für den Mandanten, aber auch für Beratungsleistungen, Schnittstellen, sonstige Software und Aufwände der Rechenzentren müssen in einem tragfähigen Verfahren aufgeteilt werden. Hierzu sind mehrere Optionen vorstellbar, die sich entweder am realen Aufwand, an der Einwohnerzahl oder an der Datenmenge des eingelieferten digitalen Archivguts orientieren. Auch die Frage, wer über Schenkungen und Deposita entscheidet und dadurch anfallende Kosten trägt, sollte geregelt werden. Insofern ist ein klares Rechte- bzw. Rollenkonzept essenziell. Es muss bei einer gemeinsamen Nutzung eines Mandanten allen Beteiligten klar sein, dass eine weitere Abgrenzung in „Submandaten“ nicht immer möglich ist. Filigran ausgestaltete Leserechte können daher nur sehr begrenzt eingeräumt werden. Die Mitarbeitenden des Verbundarchivs erhalten somit Zugriff auf das gesamte digitale Archivgut des Mandanten, ohne dass nach Institution oder Archivträger unterschieden werden könnte. Angesichts der langfristig anfallenden Kosten sollte die öffentlich-rechtliche Vereinbarung auf Dauer angelegt sein. Dennoch macht es Sinn, eine Exit-Strategie für den Fall vorzusehen, dass ein Partner den Verbund verlassen möchte. Zum einen entspricht dies gesetzlichen Vorgaben.¹⁷ Zum andern fallen u. U. für das digitale Archivgut auch nach dem Austritt weitere Kosten an, wenn Daten gar nicht oder nicht zeitnah gelöscht werden können. Es muss daher geklärt werden, wer in welchem Umfang für diese Daten aufkommt. Im Fall eines Austritts können auch die Berechnungsgrundlagen für das gemeinsam genutzte Personal hinfällig und eine Überarbeitung notwendig werden. Um das digitale Archivgut wieder nach Verbundpartnern trennen und einzeln exportieren zu können, muss es unterscheidbar sein. Für diesen Zweck sollten eindeutige IDs vergeben werden, die sich nach Verbundpartner unterscheiden und in den AIPs und den jeweiligen Verwaltungsdatenbanken nachgehalten werden. Die konsequente und korrekte Vergabe muss organisatorisch und/oder technisch sichergestellt werden, z. B. durch Pflichtfelder beim Ingest.

An dieser Stelle kann nicht auf alle vorstellbaren Varianten eingegangen werden, die im Rahmen einer Vorfelduntersuchung zu betrachten wären. Einen guten Überblick über klärungsbedürftige Punkte liefert der nestor-Kriterienkatalog für vertrauenswürdige digitale Langzeitarchive sowie die Erläuterungen zum nestor-Siegel.¹⁸ Sie können nicht nur zur Evaluation bestehender Einrichtungen verwendet werden, sondern ebenso als Checkliste bei der Gründung neuer Institutionen, ohne dass zu diesem Zeitpunkt alle Anforderungen erfüllt sein müssten. Auch die sog. NDSA-Level for digital Preservation oder das Rapid Assessment Model (RAM) der Digital Preservation Coalition könnten in vergleichbarer Weise eingesetzt werden, um regelungsbedürftige Punkte zu identifizieren, Mindeststandards sowie Prioritäten festzulegen.¹⁹ Auf diese Weise könnten viele Punkte bereits bei der Planung mitgedacht werden, um zu verhindern, dass es erst nach der Gründung der Verbundeinrichtung konfliktbehaftete Aspekte auftauchen.

Ein Verbundarchiv kann dazu beitragen, finanzielle und personelle Belastungen sowie organisatorische und fachliche Herausforderungen zu reduzieren. Jedoch können nicht alle Punkte, die sich auf die Kooperation auswirken, vorab durch eine Vereinbarung aufgefangen werden. Sollte der Wille zur Zusammenarbeit nicht von Anfang an bei allen Beteiligten vorhanden sein oder im Lauf der Zeit nachlassen, kann dies kaum regulativ gelöst werden. Dies gilt ebenso, wenn die Anforderungen bzw. Erwartungen an das Verbundarchiv vonseiten der Partner sehr stark divergieren, weil z. B. sehr heterogene Träger zusammenkommen. Eine weitere Hürde liegt gegebenenfalls in der Software-Landschaft der Verbundpartner. Sollten sehr viele unterschiedliche Softwares, Fachverfahren und DMS zur Erledigung von Fachaufgaben eingesetzt werden, führt dies zu sehr hohen Aufwänden und zu Folgefragen wie der Priorisierung von Aufgaben oder Projekten. Je einheitlicher die bei allen Verbundpartnern eingesetzte Software ist, desto besser lassen sich Übernahmekonzepte und Schnittstellen nachnutzen. In NRW wird diese Frage noch dadurch verschärft, dass es eine Reihe von regionalen Gebietsrechenzentren gibt, die z. B. unterschiedliche DMS einsetzen.²⁰ Dadurch kann die Gründung

17 Z. B. dem Gesetz über kommunale Gemeinschaftsarbeit (GkG) (wie Anm. 8) § 23, 5.

18 nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (Hrsg.), Kriterienkatalog für vertrauenswürdige digitale Langzeitarchive (Version 2). Frankfurt 2008; online unter: <https://d-nb.info/1000083241/34> (abgerufen am: 14.07.2023); nestor-Arbeitsgruppe Zertifizierung (Hrsg.), Erläuterungen zum nestor-Siegel für vertrauenswürdige digitale Langzeitarchive (Version 2.1), Frankfurt 2019; online unter: <https://d-nb.info/1189191830/34> (abgerufen am: 14.07.2023). Normenausschuss Bibliotheks- und Dokumentationswesen (NABD) (Hrsg.), DIN 31644 – Kriterien für vertrauenswürdige digitale Langzeitarchive, Berlin 2012.

19 Zu den Levels of Digital Preservation der National Digital Stewardship Alliance siehe: <https://ndsa.org/publications/levels-of-digital-preservation/> (abgerufen am 14.07.2023). Das Rapid Assessment Model der DPC findet online unter: <https://www.dpconline.org/digipres/dpc-ram> (abgerufen am: 14.07.2023).

20 Eine Übersicht über die Versorgungsgebiete der regionalen Rechenzentren findet sich unter: <https://www.kdn.de/ueber-den-kdn/portrait/> (abgerufen am: 14.07.2023).

eines Verbundarchivs über das Versorgungsgebiet eines Rechenzentrums hinaus erschwert bzw. die Auswahl möglicher Kooperationspartner eingeschränkt werden. Auch um solche Hindernisse frühzeitig zu ermitteln und auszuräumen, sollten eventuelle Partner eine Zusammenarbeit im Vorfeld umfassend und offen prüfen.

Damit die digitale Langzeitarchivierung auch und gerade von kleinen und mittleren Einrichtungen gemeistert werden kann, sind Kooperationen unumgänglich. Ansonsten besteht die reale Gefahr, dass viele kleinere und mittlere Archive die digitale Langzeitarchivierung gar nicht erst angehen, bei der Einführung oder der Umsetzung scheitern. Nicht umsonst bildete sich eine nestor-Arbeitsgruppe „Kooperation der Archive“, die Möglichkeiten und Vorschläge zusammengetragen hat.²¹ Die Verbände, die bundesweit entstanden sind, bieten hier eine wertvolle Unterstützung. Die skalierbare technische Zusammenarbeit innerhalb des DIMAG-Verbunds und die Mandantenzusammenlegung im DA NRW sind Ansätze, die digitale Langzeitarchivierung in möglichst vielen Archiven zu etablieren. Beide reagieren auf die jeweiligen Rahmenbedingungen und weisen eigene Vor- und Nachteile auf. Zusätzlich zu den Chancen, die eine Teilnahme an einer Verbundlösung mit sich bringt, zielen sie darauf ab, Kosten und Lasten für kleinere Einrichtungen weiter zu reduzieren. Interessierte Einrichtungen können aus einer ganzen Bandbreite von Optionen wählen, um die Ausgestaltung an die konkreten Anforderungen anzupassen. Die Mandantenzusammenlegung setzt voraus, dass kooperierende Archive vorab klare Vereinbarungen zu finanziellen, personellen und rechtlich-organisatorischen Aspekten geschlossen haben. Dazu gehört explizit auch eine Exit-Option und die sich daraus ergebenden Konsequenzen für die verbleibenden Akteure. Die möglichen Partner müssen sich daher bereits im Vorfeld einer intensivierten Kooperation umfangreich austauschen, zahlreiche Punkte bedenken und regeln. Auch im laufenden Betrieb werden weitere Abstimmungsprozesse und ggf. Kompromisse notwendig sein. Dies klingt nicht übermäßig verlockend. Allerdings kann es einen wichtigen Beitrag dazu leisten, die archivische Vielfalt in der Fläche zu erhalten und einem Ausbluten der archivischen Infrastruktur entgegenzuwirken. Dies nutzt langfristig der gesamten Kultursparte Archiv, da vereinzelt Leuchtturmprojekte auf Dauer nicht ausreichen werden, um die Aufmerksamkeit der Träger und sonstiger Geldgeber zu sichern.

21 Nestor Arbeitsgruppe Kooperation der Archive (Hrsg.), *Gemeinsam handeln. Vorschläge für Archive*, Frankfurt 2018; online unter: <https://d-nb.info/1152239295/34> (abgerufen am: 14.07.2023).

XDomea-Schnittstelle TIZIAN: Datenmapping inkl. AIP-Bildung

Daniel Baumann, Stadtarchiv München, Markus Schmalzl, Generaldirektion der Staatlichen Archive Bayerns

Die realisierte Aussonderungsschnittstelle für das Fachverfahren TIZIAN ist aus einem Kooperationsprojekt zwischen kommunalen und staatlichen Archivträgern hervorgegangen und dient der Archivierung der Daten eines konkreten Fachverfahrens. Der Bedarf für eine derartige Zusammenarbeit ist in doppelter Hinsicht evident. Zum einen werden immer mehr kollaborative bzw. mandantenfähige Fachverfahren eingesetzt, die nicht nur behördenübergreifend betrieben werden, sondern auch von Dienststellen unterschiedlicher Träger genutzt werden – und dies bei gemeinsamer Datenhaltung.¹ Zum anderen sind in Bayern die Landratsämter als Mischbehörden der staatlichen Verwaltung und der kommunalen Selbstverwaltung allein an die Staatliche Archivverwaltung abgabepflichtig. Dadurch fallen im staatlichen Bereich digitale Informationen zu denselben Fachaufgaben an, die in den 25 kreisfreien Städten von an das dortige Kommunalarchiv abgabepflichtigen Ämtern wahrgenommen werden. Vor diesem Hintergrund haben sich das Stadtarchiv München und die Generaldirektion der Staatlichen Archive Bayerns zu einem gemeinsamen Aussonderungsprojekt zusammengeschlossen.²

1 Fachverfahren TIZIAN

Bei TIZIAN handelt es sich um ein komplexes Fachverfahren, das im Bereich der behördlichen Veterinär-, Futtermittel- und Lebensmittelüberwachung zum Einsatz kommt. In TIZIAN werden eine Fülle von Metadaten zu bayerischen Tierzüchtern, Betrieben der Lebensmittel- und Futtermittelherstellung und Weiterverarbeitung – vom Kleinbetrieb bis zum industriellen Großunternehmen – gespeichert und bearbeitet. Außerdem werden hier auch die Detailinformationen zu behördlichen Maßnahmen wie Kontrollbesuchen, Probenentnahmen, Risikobewertungen, Beanstandungen, Ordnungsmaßnahmen etc. verwaltet. Auf dieser Grundlage ergibt sich ein feingranulares Bild der kommunalen und staatlichen Lebensmittel- und Futtermittelüberwachung sowie der Tierhaltung in Bayern in rund 840.000 Betrieben. Informationen zu diesen wichtigen Bereichen der öffentlichen Verwaltung und Daseinsfürsorge stellen bislang weitgehend ein Desiderat archivischer Überlieferungsbildung dar – kurz, die Daten sind grundsätzlich als archivwürdig einzustufen und sollten in breitem Umfang archiviert werden.

Das Fachverfahren TIZIAN basiert auf BALVI iP der Lübecker Firma BALVI GmbH einem bundesweit eingesetzten Softwaresystem zur behördlichen Überwachung im Veterinär- und Lebensmittelbereich.³ Das System ist in Bayern seit 2007 bei rund 1.500 Anwenderinnen und Anwendern in 104 kommunalen und staatlichen Behörden im Einsatz. Aufseiten des Freistaates wird das Fachverfahren im Bayerischen Landesamt für Gesundheit und Lebensmittelsicherheit, bei den Regierungen der Regierungsbezirke und bei den Landratsämtern eingesetzt, die in Bayern ebenfalls an die staatliche Archivverwaltung abgabepflichtig sind. Die Daten werden gemeinsam in einer relationalen Datenbank vorgehalten. Alle mit TIZIAN arbeitenden Behörden – ob kommunal oder staatlich – sind dabei als eigene Mandanten innerhalb des Fachverfahrens abgebildet. Entsprechend liegt die Datenverantwortung bei den einzelnen Überwachungs- und Kontrollbehörden. Verfahrenspflegestelle für Bayern ist das Bayerische Landesamt für Gesundheit und Lebensmittelsicherheit in Erlangen. Im Gegensatz zum Grundprodukt BALVI iP ist TIZIAN keine reine Datenbankanwendung, sondern ein primärdatenführendes und prozesssteuerndes Fachverfahren, mit dessen Hilfe auch definierte Dokumente der Office-Palette sowie digitale Fotos und Videos als Binary Large Objects (BLOBs) gespeichert, aus Verwaltungsprozessen erzeugt und entlang des Geschäftsgangs verarbeitet werden. Hier liegt auch der entscheidende Unterschied zu den in anderen Bundesländern eingesetzten Derivaten des Fachverfahrens BALVI iP, bei denen definierte Primärdokumente nicht direkt im Fachverfahren abgespeichert werden.

1 Vgl. Annekathrin Miegel, Aufbau eines Bewertungs- und Übernahmemanagements für Fachverfahren im Hessischen Landesarchiv. Manuskript des Vortrags für die 22. Tagung des Arbeitskreises zur Aussonderung von Unterlagen aus digitalen Systemen am 7. und 8. März 2018 in Marburg, S. 2, online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2018/strukturelle/09-Miegel_Fachverfahren.pdf. (abgerufen am: 14.07.2023); vgl. Michael Puchta, Bewertungskriterium Standardformat? Die Auswirkungen der Format- und Schnittstellenproblematik auf die Aussonderung und die Auswertbarkeit elektronischer Unterlagen im Digitalen Archiv, in: LWL-Archivamt für Westfalen (Hrsg.), Bewertung und Übernahme elektronischer Unterlagen – Business as usual? Beiträge des Expertenworkshops in Münster am 11. und 12. Juni 2013, S. 30–45, hier S. 39.

2 Vgl. die Präsentation des Projektes bei der AUdS-Tagung 2018 online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2018/praxis/19-03_Ingest_Puchta_und_Baumann.pdf (abgerufen am: 14.07.2023)

3 Vgl. Produktdatenblatt BALVI iP, online unter: <https://www.balvi.de/fachanwendung/balvi-ip2/> (abgerufen am: 14.07.2023)

2 Einsatz einer xdomea-Archivierungsschnittstelle

Die Entscheidung für die Umsetzung einer xdomea-Schnittstelle fiel aus mehreren Gründen. Zum Ersten werden bei den für die Lebensmittel- und Futtermittelüberwachung in Bayern zuständigen Stellen teils keine vollständigen teils, überhaupt keine herkömmlichen Akten mehr geführt. Vielmehr werden die wesentlichen Informationen und Dokumente rein im Fachverfahren vorgehalten. Zum Zweiten ermöglicht eine Aussonderung auf xdomea-Basis eine standardisierte Archivierung, die realistischere auch von allen betroffenen Kommunalarchiven nachvollzogen werden kann. Zum Dritten können die Daten auf diesem Weg automatisiert übernommen, ingestiert, strukturiert, erschlossen und quittiert werden, weil mit dem xdomea-Client der Staatlichen Archive Bayerns und der Scope-Schnittstelle des Stadtarchivs München entsprechende Tools zur Verfügung stehen. Die Aufwände zur Nachbearbeitung der Daten sind deshalb relativ gering.⁴ Vor allem aber entspricht die Archivierung der Daten in Akten- und Vorgangsform auch dem Entstehungskontext und Entstehungszweck der Informationen. Da uns leider die notwendigen Augurenkräfte fehlen, um Nachnutzungszwecke künftiger Generationen definieren zu können, orientieren sich die beteiligten Archive bei der Digitalen Archivierung an dem, was gesichert festgestellt werden kann: am Entstehungskontext und mithin am Provenienzprinzip.⁵

3 Konzeption der Archivierungsschnittstelle

Wie wurde nun die Schnittstelle konzipiert? Die Daten wurden vorab bewertet, sodass ein zweistufiges Verfahren gemäß xdomea-Standard umgesetzt werden konnte. Dabei setzen wir auf die bei den beteiligten Archiven zum Zeitpunkt des Projektbeginns bereits implementierte Version 2.1. Die Daten werden als ZIP-Pakete für jeweils 1000 Betriebe und mit jeweils einer großen xdomea-Nachricht nach Provenienzstellen aufgeteilt, direkt an das Anlieferungsshare der Staatlichen Archive Bayerns im landeseigenen Rechenzentrum geliefert bzw. über die Own-Cloud der Verfahrenspflegestelle für die Kommunalarchive bereitgestellt und auf demselben Weg quittiert. Wie im Ursprungssystem angelegt, werden die Daten nach Betrieben strukturiert ausgespielt und im Archivspeicher entsprechend als AIPs abgelegt, d. h. die Aktenebene bilden die Betriebe. Auf Vorgangsebene werden Vorgänge mit den wesentlichen Stammdaten des Betriebes sowie Vorgänge zu den Maßnahmen gebildet, die zu weiterer behördlicher Tätigkeit, wie etwa zu Vollzugsmaßnahmen, die Anordnung von Bußgeldern oder Kontrollen infolge anonymer Hinweise aus der Bevölkerung geführt haben. Bei Hochrisikobetrieben werden alle diese Maßnahmen der Überwachungsbehörden als Vorgänge in das zuständige Archiv übernommen. Auf Dokumentenebene werden schließlich die definierten Primärdokumente wie Fotos, Videos, Kontrollberichte, anonyme Anzeigen besorgter Bürger*innen und andere Schriftsätze archiviert. Die gezippten Abgabepakete enthalten dann neben den xdomea0503-XML-Dateien, die Primärdaten in Form von Kontrollberichten, Fotos etc. sowie Primär-XML-Dateien mit den Metadaten zu den Betriebsstätten und Maßnahmen sowie XSD-Schemadateien zu diesen XML-Dateien.

4 Projektverlauf

Auf der 22. AUdS-Tagung im März 2018 in Marburg stellte Dr. Michael Puchta (Generaldirektion der Staatlichen Archive Bayerns) die Schnittstelle aus TIZIAN in der Hoffnung vor, diese bis Ende 2018, spätestens im ersten Quartal 2019 in den Produktivbetrieb übernehmen zu können.⁶ Erreicht wurde dieses Ziel mit etwa einem Jahr Verzögerung. Die Gründe hierfür sind vielfältig.

Zunächst war die Konzeption der Schnittstelle allein in Kooperation mit der Verfahrenspflegestelle für TIZIAN, dem Bayerischen Landesamt für Gesundheit und Lebensmittelsicherheit (LGL), erarbeitet worden. Auf einem Workshop Anfang Mai 2018 musste daher die Schnittstelle noch mit dem Hersteller des Fachverfahrens, der Firma BALVI GmbH abgestimmt werden. Auf Grundlage der Workshop-Ergebnisse setzte BALVI im Anschluss einen groben Fahrplan für die Umsetzung der Schnittstelle auf, der die Produktivsetzung spätestens für das erste Halbjahr 2019 vorsah.

Als ein größeres Hindernis auf dem Weg zu einer Aussonderungsschnittstelle erwies sich die lange Zeit nicht gesicherte Finanzierung. Erst Anfang 2019 sagte das Bayerische Staatsministerium für Umwelt und Verbraucherschutz die Übernahme der Kosten zu. Bis dahin hatte die Firma BALVI ihre Arbeiten ohne Finanzierungszusage durchführen müssen. Daneben trug zur Verzögerung auch ein Entwicklerwechsel innerhalb der Firma Ende 2018 bei.

4 Vgl. Michael Puchta, Automatisierung und Standardisierung. Ein Praxisbericht aus den Staatlichen Archiven Bayerns, in: *Archivar* 3 (2021), S. 180-186.

5 Vgl. Michael Puchta, Signifikante Eigenschaften für eine „unknown community“, in: *Archivar* 3 (2020), S. 259-268, hier S. 262-268, sowie Peter Worm, Nicola Bruns, Form follows function – ein Grundsatz für die elektronische Überlieferungsbildung, in: *Aktuelle Fragen der Überlieferungsbildung*. Stuttgart 2020 (= Vorträge des 79. Südwestdeutschen Archivtags am 16. und 17. Mai 2019 in Ludwigsburg), S. 18-27, hier S. 25f.

6 Vgl. die Präsentation des Projektes bei der AUdS-Tagung 2018, online unter: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2018/praxis/19-03_Ingest__Puchta_und_Baumann.pdf (abgerufen am: 14.07.2023)

So wurden erst Ende November 2018 die ersten Testdaten ausgespielt, die aber noch einen erheblichen Anpassungsbedarf aufwiesen. Es dauerte daher ein weiteres halbes Jahr, bis die Schnittstelle schließlich soweit programmiert war, dass ausreichend Daten für Massentests zur Verfügung gestellt werden konnten.

Die Analyse der Testdaten im Sommer 2019 offenbarte dann einige Probleme. Unter anderem stellten sich plötzlich fachliche Fragen, die bei der theoretischen Konzeption der Schnittstelle noch nicht als potenziell problematisch erkannt worden waren. Die aufgetretenen Probleme und Fragen werden weiter unten im Einzelnen vorgestellt. Für das Projekt bedeuteten sie eine weitere Verzögerung, da ihre Lösung und Beantwortung einige Zeit in Anspruch nahmen.

So waren drei weitere Schleifen zu drehen, bis schließlich am 21. Januar 2020 den Staatlichen Archiven Bayerns und dem Stadtarchiv München die ersten Aussonderungspakete zur Verfügung gestellt werden konnten.

5 Erste Aussonderung und Übernahme

Das Fachverfahren TIZIAN ist, wie bereits mehrfach erwähnt, in allen kreisfreien Städten Bayerns im Einsatz. Ziel des Projekts war es daher von Anfang an, allen betroffenen bayerischen Kommunalarchiven die Übernahme und Archivierung der zur Aussonderung anstehenden Daten ihrer Behörden zu ermöglichen. Daher wurde Vertretern bayerischer Stadtarchive auf einem Workshop im September 2018 in Bamberg die xdomea-Schnittstelle aus TIZIAN vorgestellt. Die Reaktionen auf unsere Arbeit waren durchweg positiv. Dennoch wurde auch die Problematik thematisiert, wonach kleinere und mittlere Kommunalarchive im digitalen Zeitalter zunehmend gezwungen werden, ihre Bewertungshoheit aufzugeben, da ihnen die Ressourcen und die Fachkenntnisse fehlen, um an der Konzeption solcher Schnittstellen mitarbeiten zu können.

Im Januar 2019 wurden die Archive aller 24 kreisfreien Städte in Bayern kontaktiert und über die anstehende Aussonderung der zu löschenden Daten aus TIZIAN informiert. Zumindest 16 Archive meldeten sich zurück und bekundeten ihr Interesse an den TIZIAN-Daten. Seit Anfang Februar 2020 konnten sie schließlich „ihre“ Daten über die Own-Cloud des LGL für die Archivierung abholen.

Die Staatlichen Archive Bayerns und das Stadtarchiv München haben ihre TIZIAN-Daten mittlerweile in ihren digitalen Magazinen archiviert. Das Stadtarchiv München konnte über seine xdomea-Importschnittstelle die zehn ausgesonderten Pakete mit Daten zu über 9.000 Betriebsstätten übernehmen und weitgehend automatisiert zu über 9.000 Archivpaketen aufbereiten. Dabei wurde zudem in der Erschließungsdatenbank zu jedem Archivpaket eine Verzeichnungseinheit angelegt, sodass jede elektronische Archivalie mit Basisdaten nachgewiesen ist und nach ihr recherchiert werden kann.

Bei den Staatlichen Archiven Bayerns fiel die verspätete Ausspielung der Daten mit einer geplanten Versionshebung des xdomea-Clients zusammen, was die Verarbeitung bis Ende 2020 verzögerte. Mittlerweile konnten aber auch hier die Daten automatisiert ingestiert, strukturiert, erschlossen und quitiert werden.

Viele weitere Kommunalarchive haben ihre TIZIAN-Daten ebenfalls abgeholt. Für viele bayerische Kommunalarchive war dies sicherlich die erste Übernahme digitaler Unterlagen aus einem Fachverfahren.⁷

Die Stadtarchive Augsburg und Regensburg stießen allerdings bei dem Versuch, ihre Daten zu archivieren, noch auf ein weiteres Problem, welches bei der Konzeptionierung der Schnittstelle nicht bedacht worden war, nämlich die fehlende Abwärtskompatibilität des xdomea-Standards. Beide Stadtarchive arbeiten bei der digitalen Archivierung mit der Software scopeO AIS in der Version V5.3, die über eine Schnittstelle zur Übernahme von xdomea-Nachrichten gemäß der Version 2.4 verfügt. Wie sich herausstellte, lassen sich über diese Schnittstelle aber keine xdomea-Nachrichten archivieren, die gemäß einer Vorgängerversion des Standards, etwa der Version 2.1 geformt sind. Glücklicherweise war der Hersteller der Archivierungssoftware in diesem Fall bereit, sein Produkt dem xdomea-Standard und der Inkompatibilität der Versionen anzupassen, sodass scopeO AIS künftig zwei xdomea-Versionen, die Versionen 2.1 und 2.4, unterstützen wird.

Während die fehlende Abwärtskompatibilität des xdomea-Standards leider erst nach der Inbetriebnahme der Schnittstelle als potenziell problematisch erkannt wurde, konnten andere fachliche Herausforderungen glücklicherweise noch während der Testphase des Projekts identifiziert und gelöst werden.

7 Im Juli 2021 wurden die Kommunalarchive erneut angeschrieben, mit der Bitte zurückzumelden, ob sie erstens „ihre“ TIZIAN-Daten von der OwnCloud des LGL heruntergeladen haben, zweitens bei der Abholung „ihrer“ Daten auf Schwierigkeiten stießen bzw. scheiterten, und ob sie drittens die heruntergeladenen TIZIAN-Daten bereits in einem digitalen Langzeitarchiv sichern konnten. 11 Archive meldeten zurück, dass sie die TIZIAN-Pakete heruntergeladen haben, zwei Archive schrieben, ihre Verwaltungen hätten keine Daten zur Aussonderung freigegeben, während ein Archiv beim Herunterladen der Daten scheiterte. Mit Ausnahme des Stadtarchivs München hat aber bisher kein Kommunalarchiv die Abgabe in einem digitalen Archiv gesichert.

6 Archivfachliche Herausforderungen

Bei der Konzeption und Realisierung der Archivierungsschnittstelle waren verschiedene archivische Herausforderungen zu meistern. Im Fachverfahren TIZIAN werden die Daten der verschiedenen Provenienzbildner in einer gemeinsamen Datenbank nach einem Mandantenkonzept verwaltet. Hier konnten wir bei den Verhandlungen erreichen, dass die archivwürdigen Daten nicht einfach an eine Adresse ausgespielt werden, was auf Archivseite zusätzliche Aufwände nach sich gezogen hätte. Vielmehr werden die Daten durch das Bayerische Landesamt für Gesundheit und Lebensmittelsicherheit nach archivischer Zuständigkeit für die sieben betroffenen Staatsarchive auf das Anlieferungsshare im Rechenzentrum des Freistaats angeliefert bzw. für die 25 Kommunalarchive der kreisfreien Städte zur Abholung auf der Cloud der Verfahrenspflegestelle bereitgestellt. Die ZIP-Pakete der Abgabeportionen, die nach Provenienzbildnern getrennt geliefert werden, sollten außerdem neben den Primär- und Metadaten nach xdomea-Standard auch ausreichend Dokumentationsmaterial zum Fachverfahren (Benutzerhandbücher und Erläuterungen zu den XML-Tags) sowie XSD-Schemadateien zu den XML-Primärdateien für die Betriebsstätten und Maßnahmen enthalten. Für dieses Dokumentationsmaterial wird allen Abgabepaketten eine fiktive Akte mit einem ebenfalls fiktiven Vorgang „Dokumentation“ hinzugefügt. Auf diese Weise können diese Daten ebenfalls automatisiert verarbeitet werden. Außerdem gelang es, die Verfahrenspflegestelle von der Notwendigkeit zu überzeugen, dass die Inhalte zur dauerhaften Sicherstellung der Interpretierbarkeit der Informationen unverschlüsselt bzw. redundant – verschlüsselt und unverschlüsselt – übergeben werden. Die Codelisten für die zum Zeitpunkt der Übernahme verwendeten Schlüssel werden zusätzlich zu Dokumentationszwecken den Abgabeportionen beigelegt.

Als nicht ganz unproblematisch stellt sich außerdem die Datenqualität dar. Naturgemäß können nur die Daten archiviert werden, die tatsächlich im Ursprungssystem vorliegen. Wenn Attribute nicht befüllt werden, fehlen diese im archivierten Datensatz. Wenn diese Attribute für die Erschließung genutzt werden sollen, dann fehlen sie natürlich auch hier. Vor allem bei älteren Daten, die noch aus Vorgängersystemen stammen, fehlen häufiger Einzelinformationen.

7 Laufzeit der TIZIAN-Akten

Eine zentrale Herausforderung bei der Konzeptionierung der Schnittstelle war das Mapping der Datenbankinhalte aus TIZIAN in die xdomea-503-Nachricht zur Beschreibung der darin enthaltenen Akten, Vorgänge und Dokumente. Schließlich werden aus der 503-Nachricht auch die beschreibenden Metadaten der späteren AIPs und damit der zugehörigen Verzeichnungseinheiten gewonnen. Als besonders schwierig erwies sich dabei, die Laufzeit der Akten bzw. der Vorgänge zu definieren. Aber gerade die Laufzeit der Akten ist ein zentrales Metadatum, schließlich orientiert sich die Berechnung der archivischen Schutzfristen an diesem Datum.

Jede xdomea-Akte aus TIZIAN umfasst, wie oben bereits erwähnt, einen Vorgang mit Informationen zu der jeweiligen Betriebsstätte. In der Regel enthält dieser Vorgang allein einen Auszug aus der TIZIAN-Datenbank mit den Stammdaten zu der Betriebsstätte im xml-Format. In Einzelfällen aber sind dem Betrieb auch in TIZIAN gespeicherte und definierte Dokumente zugeordnet, die bei der Aussonderung ebenfalls in dem Vorgang zu der Betriebsstätte landen.

Welche Laufzeit hat nun dieser Betriebsstätten-Vorgang? In einer klassischen, dokumentenbasierten Akte bzw. einem dokumentenbasierten Vorgang definiert das älteste enthaltene Dokument den Laufzeitbeginn, das jüngste Dokument das Laufzeitende.⁸

Das Entstehungsdatum der in TIZIAN gespeicherten und einem Betrieb zugeordneten Dokumente lässt sich leicht ermitteln. Es ist in der Datenbank hinterlegt. Auch das Entstehungsdatum des Datenbankauszugs ist bekannt. Es ist das Datum der Aussonderung der Daten aus TIZIAN, denn der Auszug wird immer „on-the-fly“, während des Aussonderungsprozesses erstellt.⁹ Das Ende einer auf diese Weise berechneten Laufzeit des Betriebsstätten-Vorgangs wäre daher immer das Datum der Aussonderung der Daten aus TIZIAN.

Aber ist eine derartige Laufzeitberechnung für die vorliegenden Vorgänge sinnvoll? Für spätere Nutzer der TIZIAN-Akten wäre eine solche Laufzeit zumindest irreführend, denn die eigentliche Vorgangsbearbeitung, d. h. die Pflege der Daten der Betriebsstätte in der Datenbank, fand in der Regel viele Jahre vor der Aussonderung statt.

Wir haben es hier eben nicht mit klassischen, rein dokumentenbasierten Vorgängen zu tun, sondern mit Vorgängen, die dynamische Datenbankinhalte enthalten, die erst viele Jahre nach ihrer Bearbeitung für die Aussonderung in Form

8 Entsprechend definieren auch die „Richtlinien zur Verzeichnung von Archivgut der Staatlichen Archive Bayerns (Version 1.0)“ die Laufzeit, nämlich als Zeitraum, „aus dem die im Archivale enthaltenen Unterlagen stammen. Als Laufzeitbeginn gilt das Datum des ältesten, als Laufzeitende das Bearbeitungsdatum des jüngsten Schriftstückes.“, S.5, online unter: <https://www.gda.bayern.de/fachinformationen/erschliessung/> (abgerufen am: 14.07.2023)

9 Die „Richtlinien zur Verzeichnung von Archivgut der Staatlichen Archive Bayerns (Version 1.0)“ sehen zumindest bei der Archivierung von Datenbankschnitten die Möglichkeit vor, als Laufzeit das Jahr der Übernahme anzugeben. Siehe: Richtlinien zur Verzeichnung, wie Anm. 8, S. 72.

einer xdomea-Akte zu einem statischen Dokument geformt werden. Es erschien uns daher sinnvoller, den Vorgang nicht als eine Sammlung definierter Dokumente zu begreifen, sondern von einem materiellen Vorgangsbegriff auszugehen, der den Vorgang als Gesamtheit aller Informationen definiert, die in Bezug zu der Vorgangsbearbeitung stehen, unabhängig von der Form, in der sie gespeichert wurden.¹⁰ Entsprechend ist zur Berechnung der Laufzeit des Vorgangs der Entstehungs- und Bearbeitungszeitraum der Informationen maßgeblich, im vorliegenden Fall neben den Entstehungsdaten der in TIZIAN gespeicherten Dokumente der Bearbeitungszeitraum der im Datenbankauszug enthaltenen Informationsobjekte.

Als Laufzeit-Beginn wäre somit das Datum zu wählen, an dem der Datensatz zu der Betriebsstätte in TIZIAN angelegt wurde, und als Laufzeit-Ende das Datum, an dem der Datensatz das letzte Mal aktualisiert wurde. Beide Datumsangaben gibt es in TIZIAN. Allerdings sind es technische Daten, die vom System automatisch gesetzt werden.

Und hier liegt schon ein Teil des Problems. So ist nicht jede technische Aktualisierung des Datensatzes gleichzusetzen mit einer fachlich inhaltlichen Änderung des Datensatzes. Ist etwa eine formale Korrektur des Datensatzes, herbeigeführt durch die Ausführung eines Korrekturskriptes, eine Änderung am Datenbankinhalt, die sich bei der Berechnung der Laufzeit niederschlagen sollte?

Aber auch das Datum, das beim Anlegen des Datensatzes automatisch gesetzt wird, erweist sich bei näherem Hinsehen als problematisch. Ein Teil der TIZIAN-Datensätze wurde nämlich schon in einem Vorgängersystem angelegt und 2006 nach TIZIAN migriert. Entsprechend ist bei diesen Datensätzen das Datum der Anlage in TIZIAN das Datum ihrer Migration. Würde man daher dieses Datum für die Berechnung des Laufzeit-Beginns wählen, würde man die Bearbeitung der Datensätze in dem Vorgängersystem ignorieren.

Bleiben als Alternative noch fachliche Datumsangaben, die vom Sachbearbeiter selbst in die Datenbank eingetragen werden. Zu den Betriebsstätten sind dies die Daten der An- und Abmeldung der Betriebsstätte sowie das Datum der Risikobewertung eines Betriebs.

Die Problematik dieser Daten ist, dass sie eben nicht vom System automatisch vergeben werden, sondern von einem Sachbearbeiter, einem Menschen, der leider immer eine Fehlerquelle ist. So lag beispielsweise bei einem substanziellen Teil der Betriebsstätten in TIZIAN das Abmeldedatum vor dem eingetragenen Anmeldedatum. Das LGL musste daher ein Korrekturskript über die entsprechenden Datensätze laufen lassen, das ein fiktives Anmeldedatum eintrug und diese Anpassung als Bemerkung im Datensatz hinterlegte. Zum Teil lag auch das Datum, an dem ein Betrieb einer Risikobewertung unterzogen wurde, nach dem Abmeldedatum des Betriebs.

Neben dem Problem der Datenqualität kann man aber auch die berechtigte Frage stellen, ob die Laufzeit der Betriebsstätte mit der Laufzeit der Vorgangsbearbeitung in TIZIAN gleichgesetzt werden kann. Es ist nicht in allen Fällen davon auszugehen, dass das Datum der Anmeldung der Betriebsstätte mit dem Tag zusammenfällt, an dem der Sachbearbeiter den Datensatz zu der Betriebsstätte angelegt hat. Ganz sicher aber wird sehr häufig auch noch nach der Abmeldung einer Betriebsstätte an den Daten gearbeitet, werden diese aktualisiert, sodass das Abmeldedatum der Betriebsstätte und das Datum, an dem zuletzt an dem Datensatz gearbeitet wurde, signifikant auseinanderfallen können.

Wir haben uns dennoch bei der Berechnung der Laufzeit des Datenbankauszugs für die fachlichen Datumsangaben entschieden. Die technischen Datumsangaben hätten aus den oben genannten Gründen die Laufzeit der Vorgangsbearbeitung grob verfälscht, während die fachlichen Angaben sich der tatsächlichen Sachbearbeitung in TIZIAN zumindest annähern.

So berechnet sich die Laufzeit des Betriebsstätten-Vorgangs nun aus den Entstehungsdaten der dem Betrieb zugeordneten Dokumente sowie den fachlichen Datumsangaben des Datenbankauszugs. In der Praxis ist bei einem Großteil der Vorgänge der Laufzeitbeginn das eingetragene Anmeldedatum des Betriebs, das Laufzeitende das eingetragene Abmeldedatum.

Neben dem Vorgang mit den Daten zur Betriebsstätte umfasst eine xdomea-Akte aus TIZIAN auch Vorgänge mit Informationen zu einzelnen Tätigkeiten, wie Kontrollen oder Probeentnahmen. Diese Informationen liegen wiederum als Datenbankauszug und in Form von in TIZIAN gespeicherten Dokumenten, wie etwa Kontrollberichten, vor. Weil die Probleme bei der Berechnung der Laufzeit dieser Vorgänge vergleichbar waren zu den beschriebenen Schwierigkeiten bei der Laufzeitberechnung des Betriebsstätten-Vorgangs, wurde eine ähnliche Lösung umgesetzt.

So haben wir für die Laufzeitberechnung erneut auf das Entstehungsdatum des Datenbankauszugs verzichtet. Und auch die technischen Datumsangaben, also der Zeitpunkt, an dem der Datensatz zu der Tätigkeit angelegt wurde, und der Zeitpunkt seiner letzten Aktualisierung, wurden nicht herangezogen, weil sie sich ebenfalls aus den bereits genannten Gründen nur sehr eingeschränkt als jene Datums Grenzen interpretieren lassen, innerhalb derer tatsächlich an diesen Daten in TIZIAN gearbeitet wurde.

So greifen wir zur Laufzeitberechnung auch dieser Vorgänge auf fachliche Datumsangaben zurück. Vom Sachbearbeiter wird in TIZIAN unter anderem das Datum, an dem die Tätigkeit vor Ort in der Betriebsstätte begonnen wurde, einge-

10 Für einen materiellen Aktenbegriff vgl. Markus Schmalzl, Michael Unger, Digitales Verwaltungshandeln nachvollziehbar archivieren oder: Was ist die (E)Akte, in: *Archivar* 4 (2020), S.371–378, hier S. 374.

tragen. Auch die Daten, an denen ein Verstoß festgestellt wurde oder an dem als Konsequenz darauf eine Maßnahme festgesetzt wurde, und die Daten, an denen eine Probe versandt oder an denen der Befund der Probe vorgelegt wurden, sind fachliche Datumsangaben.

Als Laufzeit-Beginn des Vorgangs zu einer Tätigkeit wird somit das Datum gewählt, an dem mit der Tätigkeit begonnen wurde, während als Laufzeit-Ende das jüngste fachliche Datum beziehungsweise das jüngste Entstehungsdatum eines dem Vorgang zugeordneten Dokuments eingetragen wird.

Natürlich muss einem bewusst sein, dass auch hier der Tag, an dem die Tätigkeit stattfand, also z. B. eine Probe entnommen wurde oder eine Kontrolle des Betriebs stattfand, ein anderer sein kann, als der Tag, an dem die Tätigkeit schließlich vom Sachbearbeiter in der Datenbank dokumentiert wurde. Aber wie schon beim Betriebsstätten-Vorgang sind die fachlichen Datumsangaben näher an der tatsächlichen Sachbearbeitung in TIZIAN als die technischen Angaben.

Wir werden daher, bis auf eine kleinere Anpassung¹¹, bis auf Weiteres an der hier skizzierten Lösung festhalten, zumindest so lange, wie die technischen Datumsangaben nicht ein exakteres Bild der Bearbeitung der Daten in TIZIAN wiedergeben.

8 Fazit und Ausblick

Natürlich entstehen bei der Umsetzung einer Aussonderungsschnittstelle aus einem komplexen Fachverfahren hohe Kosten. Zu nennen sind zum einen die Kosten, welche die Entwicklerfirma BALVI GmbH für die Programmierung der Schnittstelle berechnete. Sie lagen in einem hohen fünfstelligen Bereich, wurden aber letztendlich vom Bayerischen Staatsministerium für Umwelt und Verbraucherschutz übernommen. Daneben aber war ein großer Personaleinsatz bei der Konzeptionierung der Schnittstelle zu leisten, sowohl von den beiden beteiligten Archiven als auch vom Bayerischen Landesamt für Gesundheit und Lebensmittelsicherheit als Verfahrenspflegestelle. Und später, in der Testphase, bedurfte es ebenfalls einer engen fachlichen und zeitlich aufwendigen Abstimmung zwischen den Archiven, dem LGL und der Entwicklungsfirma BALVI.

Als Schnittstellenkosten einzubeziehen sind zusätzlich die Aufwände, die durch die fehlende Abwärtskompatibilität des xdomea-Standards entstanden, in Form von Anpassungsleistungen an der jeweiligen Archivierungssoftware der Archive, egal, ob diese von den Archiven selbst oder, wie in Augsburg und Regensburg, von den Herstellerfirmen getragen wurden. In jedem Fall müssen Archive, das ist eine wichtige Lehre aus diesem Projekt, in Zukunft nachdrücklicher für stabile und über verschiedene Versionen kompatible Standards eintreten.

Trotz all dieser Anstrengungen kann das gemeinsame Kooperationsprojekt aber in jeder Hinsicht als Erfolg und best practice verbucht werden. Zum ersten Mal wurde in Bayern in staatlich-kommunaler Zusammenarbeit eine Archivierungsschnittstelle für ein komplexes behördenübergreifendes Fachverfahren konzipiert und produktiv gesetzt. Dabei handelt es sich für die beteiligten Archive zudem um die erste größere Übernahme nach xdomea-Standard. Im Stadtarchiv München und bei den Staatlichen Archiven Bayerns konnten mit der dort eingesetzten Software in der ersten Auspielung Daten, also Akten, zu ca. 90.000 Betrieben automatisiert ingestiert, strukturiert, erschlossen und quittiert werden. Für einige der Stadtarchive der 25 kreisfreien Städte war dies die erste Übernahme elektronischer Informationen überhaupt und damit der Einstieg in die digitale Archivierung.

Natürlich wird die Schnittstelle in den nächsten Jahren weiterentwickelt werden müssen. xdomea 3.0 wird im noch im laufenden Jahr 2021 publiziert. Das Fachverfahren TIZIAN wird voraussichtlich ab 2024 auf eine neue Version mit verändertem Datenmodell gehoben und zudem an die eAkte angebunden werden. Dies wird auch tiefe Eingriffe in die Archivierungsschnittstelle erforderlich machen. Dass damit die Fachverfahrensdaten überflüssig werden und künftig nur noch Daten aus dem eAkte-System archiviert werden müssen, ist allerdings nicht zu erwarten. Vielmehr wird hier, wie in vielen Verwaltungsbereichen, gelten, dass sich Verwaltungshandeln durch die Archivierung von Daten aus DMS/VBS alleine nicht nachvollziehbar sichern lässt, sondern lediglich durch die Übernahme von Daten aus eAkte-Systemen und angeschlossenen Fachverfahren – also von Akten in einem materiellen Sinne. Wir würden uns sehr freuen, wenn wir Ihnen in einigen Jahren dann – am Beispiel TIZIAN – eine integrative Schnittstelle vorstellen könnten, die dies leistet.

11 Bei der Berechnung der Laufzeit des Betriebsstätten-Vorgangs wird künftig das fachliche Datum der Risikobewertung ignoriert. Wie oben erwähnt, wurde auch diese Datumsangabe von Sachbearbeitern immer wieder falsch gesetzt. Eine Risikobewertung findet zwischen der An- und Abmeldung eines Betriebs statt. Ihr Datum ist daher, sofern es korrekt gesetzt wurde, für die Laufzeitberechnung des Betriebsstätten-Vorgangs unerheblich.

„Archifiltre“ – a tool of the French Social Ministries

Anne Lambertt, *Mission des Archives de France*, Chloé Moser, *Direction du numérique des Ministères Sociaux*

The paper proposed here is a feedback on a project initiated by the French Social Ministries' Records Office several years ago

1 Once upon a time...

At the beginning of this project, we started from our experiences as records managers and archivists within the ministries of Health and Labour. Every five years in France, the elections to the Assemblée Nationale trigger important archiving operations. Looking back, we noticed a major procedural change during the past several years: the proportion of electronic records, especially files and webmails, had grown significantly within both ministries. That was a challenge for us. Questions about the lifecycle of data in electronic systems have been identified quite well. Even if these questions are not resolved yet, there are now methods and solutions which are quite mature. But at the start we didn't have one for unstructured data because of the great variety in the file formats and the broad spread of data within the organizations. Nevertheless, the data within files and webmail providers are critical to document previous decisions, as was confirmed since the beginning of the health crisis resulting from the Coronavirus. Most of these decisions are prepared by editing documents, before they are exchanged by mails. Electronic document management systems do not cover the entire scope. In a context where hybrid work has emerged very quickly and the crisis has led to an emergency management, unstructured data became even more critical than before.

It is difficult to understand how information is structured without a detailed analysis of the directory. At the same time, the great number of folders, makes it quite difficult to remember where we are and where to go. We don't see the volume as a whole and it can take a long time to understand which information are included within the directory, if an appropriate file plan is applied and which files to select. For all these reasons, it is difficult to define a strategy and a methodology that can be adapted to each context and is able to convince our interlocutors to take time to implement them. If we aren't able to understand easily and quickly the meaning of what we see and the key features of the records, how could we be efficient? In the context of paper based records, what we see helps us to choose what we do. We know how much time we have and our objectives. With this, we define the level of detail in the description and the appraisal. But how to do it with electronic folders? How are we supposed to know if we have simply some sort of chaos or if we can rely that there is a form of possibly hidden organization?

Moreover, electronic records are somewhat abstract for everyone. Processing capabilities and data storage are increasing over time, and the costs per GB are decreasing so there is not as much pressure as there was at the time of paper, when archiving operations helped to give more space before a relocation or when offices were paper-filled. The issues relating to environmental costs are just beginning to emerge and the difficulties to measure them do not help to convince our interlocutors.

So most of the time, we can feel quite helpless.

That is the background that led to the birth and the development of Archifiltre. In 2017, we applied to the call for projects "Entrepreneurs d'intérêt general". This French programme allows people with digital skills to join the state administration for 10 months in order to meet challenges that intend to improve public services using digital technology and data. We have benefited greatly from the expertises of software developers and data scientists. Integrating them in the team was an exciting experience while helping us to question our practices and our challenges. At the end of these 10 months, we were authorized to continue within the "Fabrique des ministères sociaux", a division dedicated to the maintenance and continuous improvement of public agile softwares within our ministries. The tool we have created is the product of this history. Because of the way it was born, Archifiltre reflects some principles on which we have founded our project. Since the beginning of the project, we have indeed been guided by several principles.

2 Our guiding principles

First principle: be agile and user-centered

Archivists were our first users and we wanted to give them the opportunity to create new practices. The project followed its Maslow's hierarchy of needs. It is difficult to create something new when you don't know how to begin. So we wanted first to remove the "irritants". That is why our first functionality was visualization and we went on to add new functionalities from it. Each new functionality has been developed the same way. We offered some sketches and we asked our users for their advice, what allowed us to improve the tool gradually, always inspired by their returns. That is also why we chose to develop Archifiltre as an executable file. Recently, we also added a script that can be used to generate the analysis of computer servers when the volume is too big to be analysed from a personal computer. We also chose to develop a free tool whose code is open. It has enabled us to benefit from several contributions from the French National Library who developed a beta version of a METS export function.

Second principle: there are no stupid questions nor ideas, and everyone can create new opportunities

Our aim is to provide a tool that can be used in multiple ways. Any archivist should do with it whatever he or she wants. We think that the tool is there to help users to attain their objectives. We offer guidelines to use Archifiltre but we want to encourage our users to claim ownership of this tool for sharing good practices more extensively and systematically. We want them to start from their needs and show them how they could use this tool to solve their problems. We are not a software company so we experimented with different communication channels to introduce the tool and its purpose, as well as with examples of ways to use it that are based upon our own experiences. In 2020, the first lockdown due to the Coronavirus gave us the opportunity for intensive personal coaching, to try our first webinars, to improve the documentation (a wiki at Github¹), and to finalize some videos published on Youtube² (as a way of introduction). As we do them ourselves, the videos are not perfect but we hope they can be another way to discover the tool. Even if the new functionalities are more complex, we also try to maintain an easy-to-use graphic interface to offer different ways to use Archifiltre:

- an entry level for any user, including people from the staff of our organizations (secretaries or officials),
- a level for intermediary or expert use by archivists familiar with Archifiltre.

Third principle: archivists are not outdated and old practices are still useful

Electronic records, especially complex file repositories, are difficult to handle and sometimes we do not know where to begin. Nowadays, many people hope that artificial intelligence will solve their problems in the future but these technologies are expensive and require a large amount of data to be efficient. We are convinced that this quantitative approach can be complemented by a qualitative approach and that in many cases there is not enough data (or it is too heterogeneous) to enable automation without human interference. To attain this objective, we could do better if we would lose less time with some time-consuming technical operations. For example with duplicate copies of the same documents. In some cases there is no need to keep track of them but in other cases, their existence is significant. Duplicates can help to identify links between folders. Understanding of the production context, documents formats, file plans, titles and dates are criteria that are used in conjunction during the appraisal of electronic records. Appraisal remains one of our key practices and if we obtain easily the useful information we could improve our methods of appraisal of electronic records. And if we manage to develop more homogeneous practices, we could gradually become more efficient.

Fourth principle: be more responsive and more flexible

Furthermore, we often lack material and human resources, whereas the amount of electronic records is simply exploding. A first level of information is needed for a macro-appraisal, especially when we're dealing with emergencies or big volumes. After that, a detailed approach can be justified for some records. Metadata give implicit information about the records. Our conviction is that if you change the point of view and give an easy access to this information, archivists are able to choose the appropriate level of process and can adapt their practices according to the context and time available. We can even create new ways of acting. Metadata indeed change the way to describe records. If we're able to get the most resp. the most important of them, the description is more about controlling and adding new metadata in order to document the context. By doing so, we can help future research to assess the quality of information within the documents.

1 <https://github.com/SocialGouv/archifiltre-docs/wiki/Wiki-Archifiltre> (retrieved on: 14th July 2023).

2 <https://www.youtube.com/channel/UCIDwrT-Y1NY9WnEiXLMHp6w> (retrieved on: 14th July 2023).

3 The functionalities of Archifiltre

The following part introduces into the principal functionalities of Archifiltre. We have developed a version in German but there are still some parts that need to be amended. Therefore, we're open for suggestions for improvement.

Visualization in Archifiltre

To analyse a folder, the user just need to drag and drop it in the middle of the graphical user interface (GUI). The Tool works locally and there is no transfer of data if you restrain it. Moreover whatever you do doesn't affect the data, it stays in the tool. All data remains on your computer. The only data Archifiltre collects are some anonymous statistics concerning the use, but the user can change the parameters at any moment to disable this transmission. What you see after the analysis is a bar chart. Every line represents a organisational level of the file repository. For example, yellow means it's a folder. Other colours are used for documents, with different colours corresponding to different types of documents (presentations, spreadsheets...). By default, the size is proportional to the volume occupied by the folder of the file, but you can also change this parameter if you prefer to visualize by the number of documents.

The dates the documents were saved for the last time are used and are displayed. We know that this is not perfect but the Windows operating system do not offer us a better choice. This way, the oldest document is the document with the oldest saving date. For folders you can also see the median that is more satisfactory most of the time.

The title means the title of the folder or of the document. When using the search engine, it queries the titles. For security and technical reasons, we've chosen to limit the analysis to the metadata and we do not use the information of the documents themselves. But if you click on the magnifying glass, you can open the folder or the document.

Description

If the user wants to add descriptions, one can use the enrichment part of the tool. Titles can be modified. You can also add commentaries or tags if you want. In the future, we will add a functionality to help to correct the dates if needed. Blue lines allow you to see when there has been a change or a tag added. The initial title remains identified in brackets. You can also move the folders. This functionality is presently available in a beta version. The next versions of the tool will improve it significantly by fixing several bugs. You can also identify duplicates and choose which one you want to suppress. At present, duplicates are only listed at the file level but we hope to improve the functionality to use it for duplicate folders in the future.

Generate an export file

You can save your work at any moment to keep track of it or to share it with others. You just need to click the "disquette" button. The saving format is json. But, the tool offers also several others file formats for export.

Inventories

The inventory contains all the metadata and information added with Archifiltre. Every element is listed with a line for each document or folder. Each column corresponds to a type of metadata: file path, title, format. Several formats are available: csv, csv with hashes, excel and tree csv.

Preparing submission information packages (SEDA³ and METS)

Two formats of export are dedicated to the exchange with other electronic records management systems:

- A csv document that can be imported in Resip, a French tool developed by a national programme called VITAM⁴, that can produce submission information packages which comply with the second version of the SEDA standard.
- Thanks to the French National Library, a METS export is proposed in a beta version. There is work in progress to improve this functionality because METS formats can differ depending on the tools used. In the future, we wish to propose a customizable export to adapt it to the users' systems.

Audit reports

The audit report is a Word formatted document in order to be flexible to users' needs. The consensus of our users determined the type and the scope of information that is shown. It presents some metadata-based statistics: formats, duplicates, numbers of characters of the longest file path. The idea was to give the key features in one easy-to-understand document that could be used in discussions with our interlocutors.

3 SEDA (Data Exchange Standard for Archiving) is the standard used by the French archives. For more information, you can refer to the website of the French archives administration: <https://francearchives.fr/seda/> (retrieved on: 14th July 2023)

4 More information about Resip is available on the website of the Programme: <https://www.programmevitam.fr/pages/ressources/resip/> (retrieved on: 14th July 2023).

4 How to use Archifiltre: some examples

Thanks to these functionalities, Archifiltre helped us to improve our service level.

Help offices to manage their unstructured data

Archifiltre has helped us to develop an offer to the offices within the ministries because the tool enables us to deal with their unstructured data:

- Carrying a fast flash audit: we wanted to propose to the offices an overview of their share trees that could identify the warning points: a number of levels too important, old forgotten documents, a dysfunctional file plan. It helps us to define what they need to do in order to improve the situation or lose less time to find their documents. That was also more satisfactory for us. Now we have enough information to discuss with them what they want and what they are able to do, and we can track this step even if they are not available to pursue at the first try.
- Organising cleaning days: we propose organizational units to clean their share trees if they are willing to invest a little time. Archifiltre helps them to have a better overview and to determine what to keep or to delete.
- Dealing with computer servers: in 2020, we have used Archifiltre to analyse an old server from one of the ministries' directions. The initial total data volume was 950 Gb and we were able to identify what to collect and what to delete before the server had to be decommissioned.

At the moment, it's not possible to modify or suppress a folder directly from Archifiltre. You have to do it manually or using a script. But it's an option that is already identified.

Collect documents without a document management system

The French social ministries have an electronic document management system since 2021⁵. In this context, Archifiltre has helped us to improve the administration of our archival accessions:

- Generating an inventory of every deposit of records within our office. This inventory also helps us to collect information needed to document the deposit: key features like the subjects, the dates or the volume;
- Controlling the deposits: identify obvious anomalies or documents that aren't related with the main object (personal folders for example);
- Preparing for the deletion of documents: we generate a list with the documents identified by the tag "to be deleted". This list is joined to the request we send to the office that produced the records, whose agreement is required before the data are deleted.
- Giving a first analysis to help the future archival processing: a first appraisal to clarify if the documents are of interest, if their organisation is understandable. When we edit the data to send them to the National Archives, the first step of the work is done with Archifiltre before using Resip to complete some parts of the manifest (like the access rules) and generate the submission information package.

Research

Thanks to Archifiltre, we have improved the description of our electronic records. At the moment⁶, we have about 500 electronic records deposit, which represents a volume of approximately 5 terabytes. It facilitates the research we are expected to do for the officials of the ministries. Even without the information extracted from the documents, we have been able to comply with the demands. Our finding was that key words and a good knowledge of the context of production allow us to attain a good result and lose less time in finding the required information.

In conclusion, Archifiltre is still evolving and we try to continuously to improve it. In 2021, we hope to develop several new functionalities: interaction with online storage offers like Microsoft Office 365 is our priority. We would also like to develop metadata import (for exemple metadata from a document management system to improve the files metadata) and links with a thesaurus to help the use of tags. I hope this presentation will inspire you to test the tool. If you are interested to know more, please contact us for any question or presentation. I would also like to take the opportunity to make a call to volunteers. At the time, our documentation is only available in French but Github allows us to offer multilingual versions so we would be pleased to translate it in other languages with some help from volunteers.

5 It has indeed gone live in September 2021.

6 This reference refers to the volumes that had been accessed before March 2021.

Echtheitsnachweis to go.

Werkzeug zur Integritätswahrung elektronischer Reproduktionen

Tony Grochow, Konrad Meckel, Landesarchiv Thüringen

Das Projekt Digitales Magazin des Freistaats Thüringen am Landesarchiv Thüringen hat seit 2012 den Auftrag zur Vorbereitung und Einführung der elektronischen Archivierung für alle Behörden und Einrichtungen des Landes. Bei der Konzeption und Entwicklung von digitalen Archiven beschäftigt man sich – spätestens im Rahmen der Erstellung eines Sicherheitskonzepts – mit möglichen Bedrohungen. Im Rahmen einer Kooperation mit dem Lehrstuhl Mediensicherheit der Bauhaus-Universität Weimar wurden bereits frühzeitig im Projekt verschiedene Bedrohungsszenarien für das Digitale Magazin ausgearbeitet und mögliche Ansätze zum Umgang mit diesen näher betrachtet. Im Zusammenhang mit der Analyse der Szenarien zur Absicherung der Daten im Archivspeicher wurden auch randständige Fragen zur möglichen Absicherung von Nutzungspaketen aufgeworfen. Das vorliegende Werkzeug wurde in diesem Rahmen als studentisches Projekt entwickelt.¹

1 Anwendungsfall

Sofern es die datenschutzrechtlichen und archivgesetzlichen Bestimmungen in einem Benutzungsfall erlauben, können Archivbenutzer/innen Reproduktionen von Archivalien erhalten. Diese werden den Benutzer/innen vom Archiv zunehmend als digitale Nutzungspakete per Datenaustauschplattform oder Datenträger zur Verfügung gestellt (auch als „Dissemination Information Package to go“ bezeichnet, kurz *DIP2go*).

Benutzer/innen arbeiten im Anschluss mit den Reproduktionen gegebenenfalls über eine längere Zeit, bis z. B. eine Veröffentlichung von Ergebnissen oder die Abgabe einer wissenschaftlichen Arbeit erfolgt, möglicherweise liegen die vom Archiv erhaltenen Dateien auch Jahre weitgehend unbeachtet auf einem PC der Benutzer/innen. Dateien können auf nicht speziell gesicherten Ablagen verhältnismäßig einfach – absichtlich oder unabsichtlich – verändert oder gelöscht werden. Eine Überprüfung, ob die vom Archiv erhaltenen Daten auch nach Jahren noch vollständig und unverändert sind oder ob – gegebenenfalls versehentlich – Daten verändert bzw. gelöscht wurden, ist durch die Benutzer/innen allerdings nicht ohne Weiteres möglich.

Das entwickelte Werkzeug bietet den Benutzer/innen die Möglichkeit, jederzeit die ihnen vorliegenden archivischen Nutzungspakete auf Veränderungen zu überprüfen. Das Archiv kann den Benutzer/innen das Programm gemeinsam mit dem *DIP2go* oder auf anderem Weg, beispielsweise als Download, bereitstellen.

2 Konzeptioneller Hintergrund

Grundsätzlich ist eine Absicherung gegen Veränderungen an *DIP2go* und anderen Daten mit unterschiedlichen Verfahren möglich. In seinem Artikel zur Tagung des Arbeitskreises 2014 hat Prof. Stefan Lucks vom Lehrstuhl Mediensicherheit der Bauhaus-Universität Weimar mehrere kryptographische Methoden und deren Eignung für Digitale Archive betrachtet.² Digitale Unterschriften und Message Authentication Codes (MACs) schloss er u. a. wegen des erforderlichen Schlüsselmanagements für Archive als ungeeignet aus. Kryptographische Hashfunktionen dagegen wurden als grundsätzlich geeignet angesehen. Entsprechend dieser Einschätzung wurde im Rahmen des Kooperationsprojekts zwischen Bauhaus-Universität und Landesarchiv entschieden, dass das Werkzeug zur Absicherung der *DIP2go* sogenannte „Hashtrees“ oder auch „Hashbäume“ verwenden soll. Vorteil dieser Methode ist, dass nicht nur ein Schlüsselmanagement vermieden wird und alle Informationen veröffentlicht werden können, sondern auch, dass umfangreiche Nutzungspakete mit verhältnismäßig geringem Speicheraufwand abgesichert werden können.

Hashbäume basieren auf kryptographischen Hashfunktionen. Dabei handelt es sich um Funktionen, die aus einer Datei oder einer beliebigen Zeichenfolge einen Fingerabdruck, den sogenannten Hashwert, berechnen. Der berechnete Hashwert selbst benötigt nur wenige hundert Byte Speicherplatz, unabhängig von der Größe der zugehörigen Datei. Eine weitere wichtige Eigenschaft von Hashwerten ist, dass sie eindeutig ihrer Ursprungsdatei zuzuordnen sind.

Eine Textdatei die 1000-mal den Buchstaben „a“ enthält besitzt beispielsweise unter Verwendung der Hashfunktion SHA-256 den Hashwert „41EDECE42D63E8D9BF515A9BA6932E1C20CBC9F5A5D134645ADB5DB1B9737EA3“.

1 Das Werkzeug wurde 2014/2015 durch Christoph Bräutigam, damals Student der Bauhaus-Universität Weimar, entwickelt und durch Tony Grochow, Mitarbeiter im Projektteam Digitales Magazin am Landesarchiv 2021 aktualisiert und weiterentwickelt.

2 Vgl. Stefan Lucks, Digitale Magazine ohne eigenen Speicher. Wie man Integrität „fremdgespeicherter“ Archivalien sicherstellen kann, in: Jörg Filthaut (Hrsg.), Von der Übernahme zur Benutzung. Aktuelle Entwicklungen in der digitalen Archivierung, 18. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 11. und 12. März 2014 in Weimar, Weimar 2014, S. 123–127, hier S. 123f.

Ersetzt man in der Textdatei das erste „a“ durch ein „b“, ergibt sich mit derselben Hashfunktion der Hashwert „EB7F72A-09B36323AF46C121578EE51F161AA40C76DB8BD942420233A7A61DDC6“. Vergleicht man beide Hashwerte miteinander, sind Unterschiede deutlich zu sehen. Die konkrete Änderung ist aus dem Hashwert selbst jedoch nicht erkennbar.

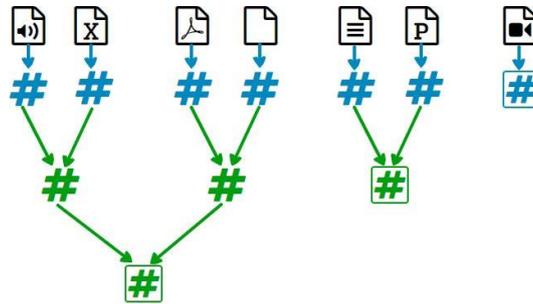


Abb. 1 Hashbaum

Zur Erzeugung der Hashbäume berechnet die Software zunächst für jede Datei eines DIP2go einen Hashwert, welche die Blätter des Baumes bilden (in Abb. 1 blau). Anschließend wird aus jeweils zwei benachbarten Hashwerten ein neuer, gemeinsamer Hashwert berechnet (in Abb. 1 grün). Es entsteht ein sogenannter binärer Baum. Die Äste eines solchen binären Baumes müssen jeweils gleich groß sein. Ist dies nicht der Fall, wird zunächst ein neuer Baum begonnen. Wenn benachbarte Bäume gleich groß sind, werden sie als Äste eines größeren Baumes zusammengefügt. Ist mehr als ein Hashbaum vorhanden, so spricht man von einem Hashwald.

3 Erstellung der Integritätsinformationen

Um die Integrität eines DIP2go außerhalb des Archivs durch Benutzer/innen überprüfen zu können, werden Informationen zum ursprünglichen Zustand benötigt. Diese sogenannten Integritätsinformationen müssen im Archiv erstellt werden, bevor sie zusammen mit dem DIP2go an die Benutzer/innen übergeben werden. Die Informationen werden durch zwei Textdateien repräsentiert. Eine Datei enthält die Hashwerte, die für die Integritätsprüfung benötigt werden und einige technische Metadaten. Die andere Datei beinhaltet die Reihenfolge, in der die Dateien des Nutzungspaketes bei der Erstellung der Integritätsinformationen eingelesen wurden. Die Integrität beider Dateien wird jeweils durch eine Prüfsumme geschützt.

Bei der Erstellung der Integritätsinformationen wird der Hashwald nach dem oben beschriebenen Verfahren erzeugt. Die exportierten Integritätsinformationen enthalten nicht den kompletten Hashwald, sondern nur eine für die Prüfung notwendige Teilmenge der Hashwerte. Es kann zwischen zwei Exportmodi gewählt werden: entweder werden erstens nur die Wurzeln oder zweitens die Blätter der Hashbäume exportiert. Werden die Wurzeln exportiert, ist nur feststellbar, dass eine Änderung vorliegt. Nur wenn alle Blätter exportiert werden, kann die Software auch anzeigen, welche Datei verändert wurde. Die weiteren technischen Metadaten umfassen die Versionsnummer des Werkzeugs, einen Zeitstempel für die Erstellung der Informationen, die ursprüngliche Anzahl an Hashbäumen und -blättern, den verwendeten Exportmodus und eine Prüfsumme.

Die Funktionalität zur Erzeugung der Integritätsinformationen wird durch eine Java-Softwarebibliothek bereitgestellt. Die Bibliothek kann so in Digitale Archive und andere Anwendungen integriert werden. Die Erzeugung der Integritätsinformationen ist zusätzlich auch über ein Kommandozeilenprogramm möglich.

4 Prüfung der Integritätsinformationen

Die Prüfung der Integritätsinformationen wird durch die Empfänger/innen des DIP2go zu einem beliebigen Zeitpunkt eigenständig durchgeführt. Dafür wurde ein minimalistisches Werkzeug als Java-Programm entwickelt.³ Die Nutzer müssen nur den Pfad des DIP2go auswählen. Das Werkzeug überprüft anschließend, ob sich die Dateien verändert haben, seit sie vom Archiv übergeben wurden. Dafür werden erst die Integritätsinformationen und anschließend die Primärdateien des Nutzungspaketes eingelesen. Das Werkzeug zeigt hierbei den jeweiligen Status der einzelnen Schritte an. Wenn es beim Lesen der Daten zu Problemen kommen sollte, beispielsweise eine Datei ist nicht vorhanden oder die Prüfsumme der Integritätsinformationen ist nicht korrekt, wird eine Fehlermeldung angezeigt (s. Abb. 2).

3 Zur Nutzung des Programms ist daher eine installierte Java-Runtime erforderlich.

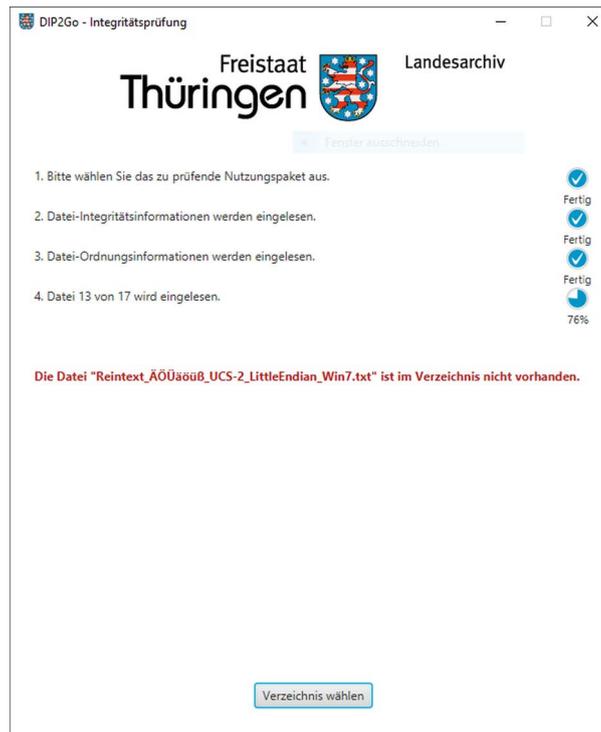


Abb. 2 Fehler beim Einlesen der Primärdateien

Zur Überprüfung wird der Hashwaid des aktuell vorliegenden Nutzungspakets berechnet. Die Primärdateien müssen hierbei in der ursprünglichen Reihenfolge eingelesen werden, da der korrekte Aufbau der Hashbäume hiervon abhängig ist. Der neu berechnete Hashwaid wird mit dem des Ursprungszustands über die Hashwerte der Baumwurzeln (eingerahmte # in Abb. 1) verglichen. Die Informationen zum ursprünglichen Hashwaid sind in den Integritätsinformationen enthalten.

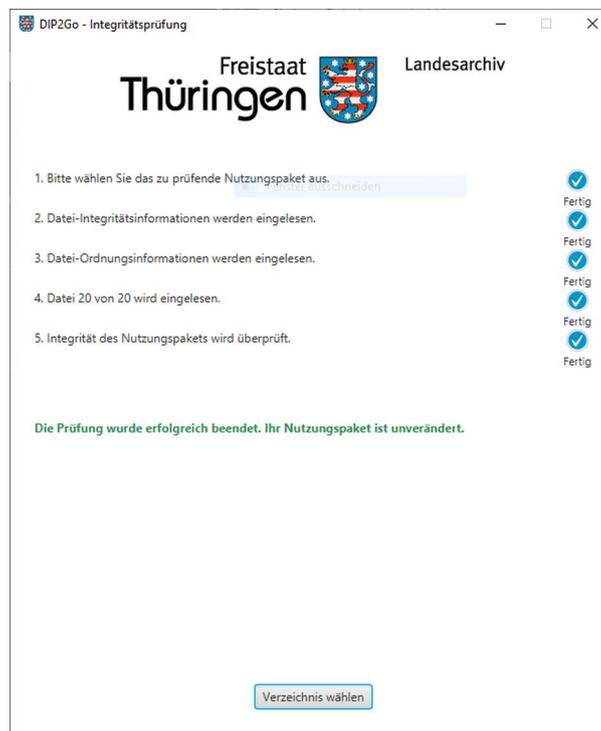


Abb. 3 Nutzungspaket ist unverändert

Sind die beiden Hashwalder identisch, so ist das Nutzungspaket seit dem Export aus dem Digitalen Magazin unverandert und es wird eine Erfolgsmeldung angezeigt (s. Abb. 3).



Abb. 4 Veranderungen im Nutzungspaket festgestellt

Werden beim Vergleich der Hashwalder Abweichungen festgestellt, so wird eine Fehlermeldung angezeigt. Wurden beim Export Hashwerte aller Blatter des Hashwaldes exportiert, kann auch genau bestimmt werden, welche Dateien sich verandert haben (s. Abb. 4).

5 Ausblick

Das Landesarchiv beabsichtigt, die Erstellung der Integritatsinformationen fur DIP2go in die Fachanwendung des Digitalen Magazins (ThELMA) zu integrieren und Archivbenutzerinnen bzw. Archivbenutzern das Prufwerkzeug zur Verfugung zu stellen. Eine Nachnutzung durch andere Archive ist ausdrucklich gewunscht.⁴ Die Oberflache des Prufwerkzeugs kann hierzu mit einem individuellen Logo und Icon des jeweiligen Archivs versehen werden.

Perspektivisch ist die Integration weiterer Funktionen geplant, z. B. Verbesserungen in der Benutzeroberflache und die Dokumentation von Prufungsergebnissen in einer Logdatei. Aktuelle Versionen des Werkzeugs werden auch weiterhin uber den GitHub-Account des Landesarchivs veroffentlicht. Hinweise, Anregungen, Wunsche und eine Beteiligung an der Weiterentwicklung der Software sind ausdrucklich willkommen.

4 Das Werkzeug steht unter der Open-Source-Lizenz LGPL und steht im Github-Account des Landesarchivs Thuringen zum Download und zur Nachnutzung kostenfrei zur Verfugung: https://github.com/Landesarchiv-Thuringen/dip_integrity (zuletzt abgerufen am: 14.07.2023)

