

10001000100011110100101
10110101101000011101101
0123.0TAGUNG1DES1110010
10111ARBEITSKREISES1001
10110011011011001011010
01010ARCHIVIERUNG01
1101110VON110000000001
010111UNTERLAGEN110111
10111011AUS110000011100
110110DIGITALEN1100101
111001SYSTEMEN0010111
1101100011000000100
110012.0UND013.0MÄRZ11111
001101020191PRAG011
00101001010010100111
10011111NATIONALARCHIV101
100101101101100110011
100110000  100111100
101000100 National Archives 111001101
Czech Republic
101001100110100110110

110101101000011101101
010ARCHIVIERUNG01
0110VON110000000001
1011UNTERLAGEN11011
111011AUS110000011100
1101DIGITALEN1100101
100SYSTEMEN0010111
0010101PRAHA20190
0010100101001010011

Editor: Karolína Šimůnková, Milan Vojáček

Kolektivní monografie vznikla na základě konference Archivierung von Unterlagen aus digitalen Systemen (Archivace záznamů z digitálních systémů) a souvisejících seminářů, které se konaly ve dnech 12. a 13. března 2019 v Národním archivu v Praze.

© Národní archiv, 2019
www.nacr.cz

ISBN 978-80-7469-100-3 (pdf)

23. Tagung des Arbeitskreises

ARCHIVIERUNG VON UNTERLAGEN AUS DIGITALEN SYSTEMEN

12. und 13. März 2019, Nationalarchiv Prag

Inhaltverzeichnis

LEITIDEEN, PROJEKTE, NETZWERKE

Ein Überblick zur Archivierung digitaler Unterlagen in Deutschland	9
Kai NAUMANN (Landesarchiv Baden-Württemberg)	
Der Stand und die Perspektiven der digitalen Archivierung in der Tschechischen Republik	16
Milan VOJÁČEK (Nationalarchiv Prag)	
Das Projekt DIMAG. Sachstand 2019	21
Christian KEITEL (Landesarchiv Baden-Württemberg)	

BEWERTUNG UND ÜBERNAHME

Qualitative Bewertung im Hinblick auf Informationsgehalt, Verwendung und Kostenmetrik Juni 2019	35
Martin KAISER (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen – KOST)	
Übernahme von landwirtschaftlichen Daten aus Fachapplikationen	41
Bernhard STÜSSI, Flurina CAMENISCH (Staatsarchiv Graubünden)	
DCAT-AP.de. Der Metadatenstandard zum Austausch offener Verwaltungsdaten	53
Paul FLAMME (Staatsarchiv Hamburg)	

ERHALTUNGSSTRATEGIEN UND ERHALTUNGSMAßAHMEN

Alles Blockchain oder was? Nutzen und Grenzen der Distributed-Ledger-Technologie in der digitalen Archivierung	67
Björn STEFFENHAGEN (Landesarchiv Sachsen-Anhalt)	
Vom Umgang mit fehlerhaften Daten im Ingest. Ein Werkstattbericht aus der Entwicklung der DA-NRW Software Suite (DNS).	81
Jens PETERS (LVR-InfoKom)	

ERSCHLIEßUNG UND NUTZUNG

„Ich packe mein AIP und nehme mit...“ Dateistrukturierung im Digitalen Magazin des Landesarchivs Thüringen (ThELMA) 91

Svenia POHLKAMP, Konrad MECKEL (Landesarchiv Thüringen)

Von der Archivvorstufe zum Archiv: (Forschungs-)Datenmanagement im Kontext von Gestaltung und Kunst 99

Jürgen ENGE, Tabea LURK (Hochschule für Gestaltung und Kunst FHNW Basel)

Die Indexierung von audiovisuellen Archiven unter Benutzung von Methoden der automatischen Sprach- und Bilderkennung 120

Pavel IRCING, Jan ŠVEC (Westböhmische Universität in Pilsen)

Service Design für den digitalen Lesesaal. Ein Werkstattbericht der Staatsarchive Basel-Stadt und St. Gallen 129

Esther BAUR – Lambert KANSY (Staatsarchiv Basel-Stadt), Martin LÜTHI (Staatsarchiv St. Gallen)

AUDS-CAMP

WIE GEHEN WIR MIT DIGITALEN MEDIEN ALS BEILAGE EINER PAPIERNEN AKTE UM?

Wie gehen wir mit digitalen Medien als Beilage einer papiernen Akte um? 143

Zusammenfassung: Kai NAUMANN (Landesarchiv Baden-Württemberg)

ELEKTRONISCH SEI DIE AKTE, HILFREICH UND GUT. BERATUNG VON BEHÖRDEN BEIM UMSTIEG AUF DIE E-AKTE

Elektronisch sei die Akte, hilfreich und gut. Beratung von Behörden beim Umstieg auf die E-Akte 147

Zusammenfassung: Christine FRIEDERICH, Martin SCHLEMMER (Landesarchiv Nordrhein-Westfalen)

MIT BPMN UND REST ZU EINER KLAREN UND OAIS-REFERENZMODELL-KONFORMEN SOFTWARE-ARCHITEKTUR

Mit Bpmn und Rest zu einer klaren und Oais-Referenzmodell-Konformen Software-Architektur 153

Zusammenfassung: Frank OBERMEIT (Landesarchiv Sachsen-Anhalt)

Mit BPMN und REST zu einer OAIS-konformen Software-Architektur 155

Frank OBERMEIT

VOLKSZÄHLUNG 2021 UND DIE ROLLE DER DIGITALEN ARCHIVE

Archivcamp „Volkszählung 2021 und Rolle der digitalen Archive“ 167

Zusammenfassung: Christian KEITEL (Landesarchiv Baden-Württemberg)

Statistik im Archiv – Eine schwierige Beziehung	169
Christian KEITEL (Landesarchiv Baden-Württemberg)	
Zur Problematik der elektronischen Archivierung von Volkszählungen 2011 und 2021 in der Tschechischen Republik	177
Karolína ŠIMŮNKOVÁ, Martin ŠISLER (Nationalarchiv Prag)	
VALIDIERST DU NOCH ODER ARCHIVIERST DU SCHON?	
Validierst du noch oder archivierst du schon? Der Bedarf eines Workflow-Managements in der Formatverifikation	187
Zusammenfassung: Christine TRÄGER, Daniel WITTMANN (Landesarchiv Thüringen)	
ARCHIVSTANDARDS UND BEST PRACTICES	
Archivstandards und Best Practices	201
Zusammenfassung: Christine GIGLER (Archiv der Erzdiözese Salzburg)	
Komplexe Realitäten – Komplexe Metadaten. Wie man Standards, Best Practices und Realität zusammenführt	202
Marta RIESS (Internationale Atomenergie-Organisation)	
Aktuelle Standards und Normen in Records Management und beweisbarer Langzeitspeicherung	210
Tomasz KUSBUR (Fraunhofer FOKUS), Ulrike KORTE (Bundesamt für Sicherheit in der Informationstechnik), Steffen SCHWALM (msg systems)	
Die nestor-AG „Kooperation der Archive“ und der nestor-Archivstandard	233
Christine GIGLER (Archiv der Erzdiözese Salzburg)	
DIGITAL OBJECTS AND THEIR PRESERVATION IN ACADEMIC WORLD	
Digital Objects and their Preservation in Academic World	241
Zusammenfassung: Zdeněk VAŠEK (Karls-Universität in Prag)	
Autorinnen und Autoren	243

LEITIDEEN,
PROJEKTE,
NETZWERKE

EIN ÜBERBLICK ZUR ARCHIVIERUNG DIGITALER UNTERLAGEN IN DEUTSCHLAND

Kai Naumann

Wenn wir heute Leute auf der Straße nach digitaler Archivierung fragen, dann steht vor ihrem geistigen Auge eine Auswahl digitaler Speichermedien oder ein Speicherdienst wie Dropbox oder Google Docs. Nicht allen Befragten wird bekannt sein, dass sich digitale Objekte, sobald man sie vernachlässigt, schnell in Nichts auflösen. Digitale Archivierung ist also Perspektivsache, wie in Tabelle 1 zu sehen ist. Dieser Text wird sich im Folgenden nur der Perspektive der deutschen Archive im Sinne der Archivgesetze widmen, aber nicht ohne Nachbargebiete mit in die Überlegungen aufzunehmen.

Betrachter	Zeitliches Ziel	Digitale Archivierung
Bürger, Verbraucher, Kleinbetriebe	10 bis 20 Jahre, persönliche Dinge länger	Gewinnt an Relevanz
Großbetriebe, Industrie, Wissenschaft	10 bis 20 Jahre, selten bis 80 Jahre	Überwiegend der Normalfall
Öffentliche Verwaltung, Gerichte	10 bis 30 Jahre, seltener >100 Jahre	Noch eine Ausnahme
Archive i. S. Archivgesetze	auf Dauer	Teils Normalfall, teils Ausnahme

Tab. 1: Digitale Archivierung als Perspektivsache.

In der Domäne der Archive ist der anspruchsvolle Auftrag klar, jedoch ist das Ausmaß der Umsetzung noch sehr vielfältig. Große Organisationen wie die Landesarchive sind bereits im Produktivbetrieb, die mittleren machen sich gerade auf den Weg. Doch diejenigen Archive mit ein bis vier Mitarbeitern, die unseren Beruf in der Fläche präsent machen, sind noch unzureichend beteiligt – zumindest, wenn wir die bundesweite Perspektive einnehmen.

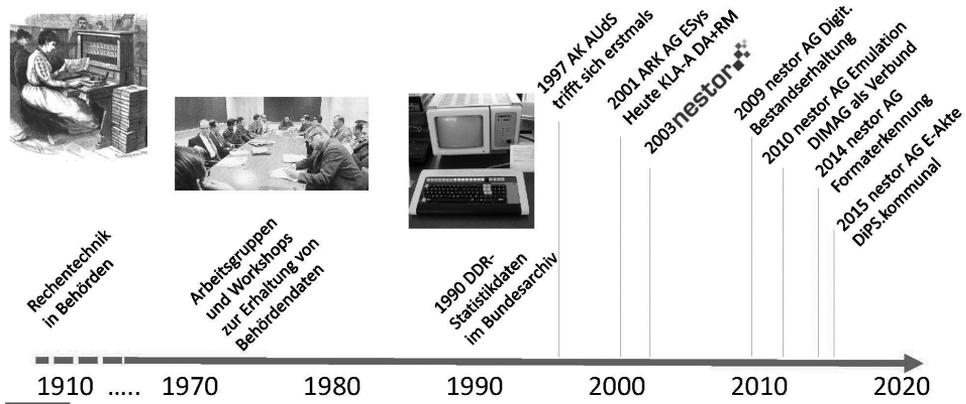


Abb. 1: Archivierung digitaler Unterlagen in Deutschland 1910–2019. Grafik: Verfasser.

Um unseren heutigen Stand darzustellen, ist ein Rückblick auf die letzten Jahrzehnte sinnvoll. Wir sehen in Abb. 2, dass die Beschäftigung mit der Aufgabe in den Archiven zunächst theoretisch geblieben ist, obwohl seit über 100 Jahren in Deutschland maschinenlesbare Unterlagen anfallen. Diese Theorie lastigkeit hängt eng mit dem Habitus der Archiv- und Geschichtswissenschaft zusammen, der die akademische Veranstaltung und die Veröffentlichung in Buchform gegenüber sonstigen produktiven Ergebnissen häufig überbewertet. Doch wir haben seit Ende der nuller Jahre endlich den Umschwung von der Theorie zur Praxis geschafft, was sich in der Gründung der heute maßgeblichen Digitalen Archivverbände und einer Vielzahl von Gremien geäußert hat.

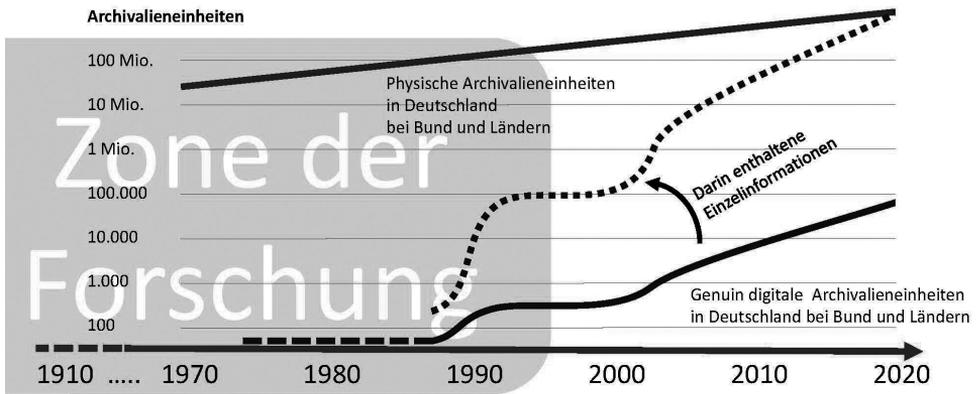


Abb. 2: Wieviel digitale Unterlagen wurden bisher bei Bund und Ländern archiviert? Eine Schätzung. Grafik: Verfasser.

Diese Wende zeigt sich auch in Zahlen. Die archivistische Statistik über digitale Unterlagen steckt noch in den Kinderschuhen, denn was für den einen eine Datenbank ist, ist für den anderen eine Serie von tausend Datenobjekten und für den nächsten eine multimediale Sammlung. Das in Abb. 2 gezeigte Diagramm über die staatlichen Archive in Bund und Ländern ist daher als persönliche Schätzung und nicht als offizielles Zahlenwerk zu verstehen. Es verwendet eine logarithmische Skala. Meine Schätzung ist, dass sich die Zahl der physischen Archivalieneinheiten in den letzten 50 Jahren von der Größenordnung

von 40 Millionen in die Richtung von einer Milliarde entwickelt hat. Wohlgermerkt sind hier nicht Akten in den Gerichten und Behörden gemeint, sondern nur die Bestände der staatlichen Archive. Über viele Jahre hinweg spielten digitale Objekte in dem Zuwachs keine Rolle. Zweimal wurde das Bundesarchiv aktiv: einmal in den 1970er Jahren, dann fünfzehn Jahre später mit einem digitalen Erbe der Deutschen Demokratischen Republik, und zwar den Statistikdaten der Zentralbehörden. Für mehr fehlte aber wiederum die Motivation, und erst ab Mitte der 2000er Jahre erblicken wir – endlich unter Beteiligung der Landesarchive – einen stetigen Anstieg der Zahlen.

Dieser Anstieg wird dann besonders eindrucksvoll, wenn wir (gestrichelt) die Archivalieneinheiten aufteilen in individuelle Informationseinheiten. Diese Einzelinformationen, die vor allem in Datentabellen und statistischen Mikrodaten stecken, sind inzwischen schon so zahlreich in unseren Beständen, dass wir aus dieser Warte binnen weniger Jahre eine Parität zwischen der Anzahl papierner und der Anzahl digitaler Objekte erreicht haben werden. Es ist aber nicht verwunderlich, dass diese an sich ganz attraktive Nachricht kaum in unserer Öffentlichkeitsarbeit verwendet wird, denn die Archivgesetze sehen für die neueste Schicht der Überlieferung eben noch keine Nutzung vor. Was im Diagramm als Zone der Forschung bezeichnet ist, das endet zur Zeit regelmäßig in den späten 1980er, in manchen Ländern in den 1990er Jahren.

In jedem Fall können wir aber unseren Trägern bei Bund und Ländern, in Kommunen und sonstigen Einrichtungen augenfällig demonstrieren, dass dem investierten Geld ein reales Wissenskapital gegenübersteht, das die Forschung der Zukunft beflügeln wird. Diese Erkenntnis ist in meinen Augen der Grund für den Erfolg der digitalen Archivierungsverbünde, der an Abb. 3 abzulesen ist. Das Diagramm ist Ende letzten Jahres entstanden, und wie dem AUdS-Tagungsprogramm zu entnehmen ist, dürfen wir ScopeOAIIS inzwischen als offizielles Produkt für die bayerischen Kommunen einordnen. Einige Bundesländer und unsere Nachbarn im Süden sind, wie der Karte zu entnehmen ist, noch in Verhandlungen.

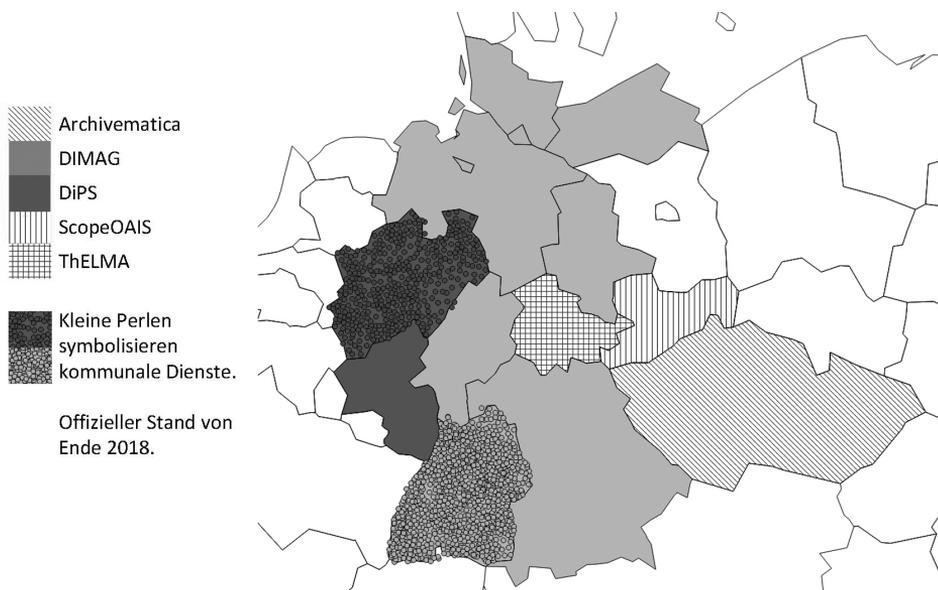


Abb. 3: Digitale Archivsysteme bei staatlichen und kommunalen Archiven. Deutsche Bundesländer und Tschechische Republik. Grafik: Verfasser.

Was in einer Übersicht wie dieser sicherlich noch erwähnt werden muss, sind die Archive der Wissenschaft, der Medien, der Kirchen sowie die großen Forschungsbibliotheken und Informatik Institute, die oft mit Rechenzentren zusammenarbeiten. In Baden-Württemberg und Hessen wird DIMAG für die Archive der Hochschulen angeboten. Beim Archiv der Max-Planck-Gesellschaft in Berlin laufen Bemühungen, die Archivierung von E-Mail-Konten zu ermöglichen. Für die Wirtschaftsarchive kümmert sich der Arbeitskreis Elektronische Archivierung bei der Vereinigung deutscher Wirtschaftsarchivare e. V. um die Verbreitung von Wissen und Lösungsansätzen zum Thema. Die Archive der Medien haben sich in den letzten Jahren intensiv großskaligen Speicherlösungen, Change Management und Machine Learning auseinandergesetzt.

In den verwandten Bereichen, die ebenfalls eine digitale Archivierung auf Dauer gewährleisten müssen, entstehen ähnliche Verbünde wie bei den Archiven. So gibt es die Anwender der israelischen Softwarelösung ExLibris Rosetta, die gemeinsam von den internationalen Netzwerken um ExLibris profitieren. Dazu gehören die Bayerische Staatsbibliothek München, die Technische Informationsbibliothek Hannover, die Zentralbibliothek für Medizin Köln und die Zentralbibliothek für Wirtschaftswissenschaften Kiel. Das Literaturarchiv Marbach, die Anna-Amalia-Bibliothek Weimar und die Herzog-August-Bibliothek Wolfenbüttel treiben die Entwicklung digitaler Dienste gemeinsam voran. Ein Ergebnis der Marbacher Forschung, der 2013 von Jürgen Enge entwickelte Indexer, wird im Rahmen des AÜS 2019 erneut erwähnt – diesmal in einer neuen Mission. Auch die GESIS (Leibniz-Institut für Sozialwissenschaften) und die Forschungsdatenzentren des Statistischen Verbunds (Bund und Länder) arbeiten daran, ihre Mikrodaten dauerhaft verfügbar zu machen.

Obwohl Rechenzentren für Archive eher die gleiche Rolle spielen wie ein Architekturbüro für physischen Magazinraum, haben auch diese unter Umständen eigene Projekte, die uns beim Bewältigen der Aufgabe helfen. Darunter fallen das Vorhaben digiS des Zuse Institut Berlin, das auf Archivemata setzt, und das bwFLA Projekt beim Rechenzentrum der Universität Freiburg im Breisgau. In Freiburg wurde eine automatisierte Bereitstellung veralteter Software und Betriebssystemen entwickelt, die international Anerkennung gefunden hat und die in der Nachfolge des EMIL-Projekt der DNB inzwischen auch die praktische Anwendungsreife erlangt hat.

Ein neues Spielfeld in diesem Bereich, dem eine erhebliche Bedeutung zukommen wird, sind die Forschungsdatenprogramme, die derzeit im Fokus der Wissenschaftsförderung stehen und erhebliche Möglichkeiten zur Akquise von Fördermitteln bieten werden. Die neu entstandenen Internetseiten forschungsdaten.info und forschungsdaten.org vermitteln entsprechende Informationen.

Die Situation bei der Nachwuchsrekrutierung zeigt Abb. 4. Hier sind die klassischen Ausbildungsstätten Marburg und Potsdam zu sehen, aber es sind auch die Hochschule Köln, die Stuttgarter Hochschulen und die Hochschule Darmstadt erwähnt. Es wird sich meines Erachtens nicht als falsch erweisen, wenn wir auch die vielen anderen Informatiklehrstühle in den sonstigen Hochschulen aufsuchen, um unsere Anliegen vorzubringen – dazu weiter unten noch mehr.



Abb. 4: Ausbildungsstätten für archivarisches Fachkompetenzen (Hochschul- und Universitätsniveau) in Deutschland (große Punkte), ergänzt durch Hochschulen mit Informatikangebot (kleine Punkte). Grafik: Verfasser.

Insgesamt können wir in Deutschland die Lage für unsere Mission als günstig einstufen. Digitale Archivierung ist inzwischen ein Thema, das auch der Mensch auf der Straße versteht. Wir haben das Gehör unserer Träger und der abgebenden Stellen. Hieraus haben sich für uns einige Positionen der Stärke ergeben:

- Wir haben die Vision von digitalen Archivierungsverbänden verwirklicht. Mit DiPS in Nordrhein-Westfalen, mit DIMAG in Baden-Württemberg, mit den Rosetta-Anwendern ist eine Vision Wirklichkeit geworden, die 2009 noch unbekannt war. Die Verbände sorgen nicht nur für technische Infrastruktur, sondern vor allem für ein breites internes Wissensmanagement und für die nötige Schulung.
- Wir haben die Möglichkeit, gemeinsam Normen zu schaffen. Auch dazu gibt es im Rahmen des AUdS 2019 Neuigkeiten.
- Wir sind sehr gut untereinander vernetzt (z. B. AUdS, Nestor-AG Kooperation der Archive, KLA-Ausschüsse Digitale Archive und Records Management). Hier möchte ich auch die neuerdings erschienenen Bücher erwähnen: Tre-

vor Owens mit Digital Preservation as Craft and Science, Christian Keitel mit 12 Wegen ins Archiv und der Band von Michael Puchta und mir über Kreative digitale Ablagen.

- Wir denken aus der zeitlichen Vogelperspektive. Als Historikerinnen und Historiker argumentieren wir weitsichtig und unterscheiden zwischen Modebegriffen und langfristig zielführenden Faktoren.
- Unsere Träger können nicht auf uns verzichten, die Budgets sind stabil.
- Die Wissenschaftsförderung bietet uns Forschungs- und Entwicklungsgelder. Gerade die Hinwendung der Wissenschaftsförderung zum neuen Thema Forschungsdatenmanagement kann uns unsere Aufgaben erleichtern, wenn wir uns um Forschungsgelder bemühen.

Doch seien wir selbstkritisch genug, auch unsere Defizite hinreichend zu kennen. Diese sind nach meiner unmaßgeblichen Meinung folgende:

- Wir denken genug an Richtlinien, aber wir können noch mehr an Umsetzung denken. Die meisten von uns sind von der Ausbildung her Historiker. Im Gleichklang mit der Rechtswissenschaft sind wir dazu ausgebildet, Texte zu ermitteln, sie zu deuten und die Deutung zu vertextlichen. Die Aufgabe der digitalen Archivierung geht eher in die Arbeitsweise der Informatik: wir müssen Bedarfe ermitteln, Abläufe gestalten, Daten übernehmen und nutzungsgerecht umwandeln. Nicht die Texttreue ist unsere Kardinaldisziplin, sondern die Treue zum Nutzer.
- Wir brauchen mehr Kontakt zu Disziplinen außerhalb von Geschichte und politischer Bildung. Diese Diagnose ist ein wenig harsch, denn viele Kolleginnen und Kollegen gehen ohne Scheuklappen in die unterschiedlichsten Bereiche von Wissenschaft, Wirtschaft und Verwaltung, sie sei aber zur Diskussion gestellt.
- Unsere Träger müssen den Investitionsbedarf für unser neues Geschäft noch besser in den Blick bekommen. Es hängt aber an uns, die Träger von der Notwendigkeit zu überzeugen, die digitale Archivierung auf Dauer zu ermöglichen.
- Wir üben zu wenig den Umgang mit Technik. Suchen wir also aktiv nach Aus- und Fortbildungsangeboten, die mit Datenbank allgemein, mit Excel und Access, ja selbst mit Programmieren zu tun haben.
- Die Informatiker finden in der Erhaltung digitaler Daten laut Deutscher Gesellschaft für Informatik eine „Grand Challenge“. Wir haben noch zu wenige Informatiker von dieser Grand Challenge begeistert. Diese Begeisterung sollten wir einfach hervorrufen, die Gelegenheit ist günstig.

Beim Workshop „Von der Theorie zur Praxis“ zur Bestandserhaltung im Bundesarchiv Koblenz im November 2018 hat William Kilbride (Digital Preservation Coalition Glasgow/York) die Situation gut zusammengefasst:

- Wir reden über undeutliche Gefahren für Daten oder Software, aber die Verluste sind echt und betreffen uns alle im Alltagsleben.
- Diese ist eine soziale und technische Herausforderung. Wenn wir nur technische Lösungen anbieten, haben wir keinen Erfolg.
- Digitale Bestandserhaltung handelt nicht von einem „dunklen Zeitalter“.
- Es geht darum, dem Versprechen der digitalen Welt gerecht zu werden.

So düster die Aufgabe der digitalen Archivierung in ihrer ganzen Komplexität und Mächtigkeit auch vor uns stehen mag – wir haben im letzten Jahrzehnt eine dreifache Wende geschafft. Erstens: Wir haben bereits digitales Archivgut

in großen Mengen übernommen. Zweitens: wir archivieren nicht mehr allein, sondern im Verbund. Drittens: Wir müssen uns nicht mehr fremden und ungeprüften Richtlinien unterwerfen, sondern wir können gemeinsam an praxisnahen Normen arbeiten. Die übrigen Hindernisse werden früher oder später auch noch überwunden.

Summary

To archivists in Germany, the task of archiving digital objects is clear, but the extent of its deployment is still quite sparse. To big players like some state archives, the task is an everyday routine, but midsized and smaller archives are only just preparing to enter the digital domain. During this decade, the community has moved on from theory to practice by establishing consortia for software development (DIMAG, DiPS, ThELMA, ScopeOAI, Archivemata, ExLibris) and working groups for standards and education (Nestor, KLA). A quick glance on still unofficial statistics shows that the number of digital objects in the custody of public archives is already considerable and serves as a valuable source for future historians. German universities are investing in research and education in this area. The Deutsche Forschungsgemeinschaft and others academic funders are investing heavily on preservation of research data.

If there are any shortcomings in the German digital archives community, those may be: an excessive reliance on standards and too little attention on feasibility, a certain lack of everyday practice with IT and programming, and a reluctance to bargain for extra funding. Also, we could do more to inspire computer scientists to work in this area.

The complex and daunting task of preserving digital content might still evoke gloomy feelings, but we have acquired mastery over big issues in three ways. First: we have already appraised and transferred large numbers of digital-born material. Second: we don't work alone but in consortia and also with citizens' networks. Third: we don't have to rely on standards made by others but we can work out definitions and specifications that fit our needs. The other obstacles will be overcome sooner or later, too.

DER STAND UND DIE PERSPEKTIVEN DER DIGITALEN ARCHIVIERUNG IN DER TSCHECHISCHEN REPUBLIK

Milan Vojáček

Bevor ich zum Stand und zu den Perspektiven komme, muss ich das Thema ein bisschen thematisch begrenzen. Was ich in diesem Beitrag unter dem Stichwort digitale Archivierung verstehe? Es handelt sich um Langzeiterhaltung der elektronischen/digitalen Informationen im Sinne der OAIS Referenzmodell im Bereich der Gedächtnisinstitutionen. Langzeiterhaltung gehört zu den grundsätzlichen Aufgaben von Archiven, Bibliotheken und Museen. Seit mindestens 30 Jahren kann man Interessen an der Langzeiterhaltung der digitalen Gegenstände in der Tschechischen Republik beobachten.

Digitales Archiv des Nationalarchivs als Zentrallösung für Archivgesamtnetz

Die ersten übernommenen digitalen Dokumente im Nationalarchiv waren im Jahr 1997 Briefe vom Philosophieprofessor Erazim Kohák. Im Jahr 2001 wurde vom Nationalarchiv die erste systematische Untersuchung der elektronischen Dokumente in den Zentralen Behörden durchgeführt. Untersuchungsergebnisse haben starke Impulse zur Notwendigkeitslösung dieser Problematik gegeben. Im Jahr 2008 wurde ein technologisches Projekt für den Digitalarchivaufbau publiziert. Es hat sich damals um die erste grundlegende Analyse der digitalen Archivierung im Tschechischen gehandelt. Es sollte ein digitales Archiv des Nationalarchivs als Zentrallösung für Archivgesamtnetz entstehen. Das so genannte National Digitalarchiv (Národní digitální archiv) wurde als Zentraldienst der Gewährleistung der Langzeitarchivierung von digitalgenuin Archivalien konzipiert. Schon hier wurden digitale Akten im Zusammenhang mit E-Government akzentuiert. (Die Bewertung und Übernahme von Behörden DMS/VBS.)

Nach der komplizierten Erfindung der Projektfinanzierung wurde in den Jahren 2011 bis 2014 das Projekt *National Digitalarchiv* realisiert. Während der Projektrealisierung haben Probleme des prozeduralen Charakters wegen der Rivalität von Herstellern DMS/VBS in der Tschechischen Republik entstanden. Infolge war es unmöglich die öffentliche Ausschreibung für die Lieferung des Digitalarchivs zu beenden. Das Nationalarchiv musste seine Strategie ändern und eine Notfalllösung sichern. Es gab damals zwei Varianten. Die erste war die Anknüpfung zu DIMAG. Die Kollegen aus Baden-Württemberg waren vorbereitet das Nationalarchiv als den Nutzer von DIMAG unter günstigen Bedingungen zu willkommen. Obwohl diese Variante vom Nationalarchiv nicht genutzt wurde,

haben wir gute Kontakte angeknüpft. Damals waren wir auch zum ersten Mal bei AUDS 2014 in Weimar.

Die andere Variante, die schließlich realisiert wurde, war eine eigene Lösung. Als Kern des Digitalarchivs wurde das kanadische Open Source Archivemata gewählt. Seit dem 1. November 2014 gab es Testbetrieb, seit dem 1. Januar 2016 normaler Betrieb des digitalen Archives. Das Nationalarchiv und Regionalarchive konnten Bewertung und Übernahme von digitalen Akten aus DMS/VBS der Behörden laut des Archivgesetzes (seit 2009 waren DMS/VBS pflichtig) realisieren.

Im Jahr 2012 wurde das Archivgesetz grundsätzlich geändert, es gab vor allem im Bereich des Archivportals mehrere neue Pflichten. Es sollte das Zentralportal für Archive entstehen, das nicht nur Bewertung und Übernahme der digitalen Dokumente ermöglicht, sondern auch Metadaten von Archiven, Archivalien und Behörden und anderen Urhebern sammelt und veröffentlicht. Im Hinblick auf Finanzeinschränkungen wurden diese Pflichten im neugebauten Nationalarchivportal nur teilweise gesichert. Das Portal wurde im Jahr 2017 in Betrieb gestellt. Man muss viele Archive für die Arbeit mit dem Portal einschulen (Bewertung, Übernahme, Beschreibung der Behörden etc.)

Es ist gelungen neue Finanzierung für das Projekt *National Digitalarchiv II* zu bekommen, das seit 2017 realisiert ist und im Dezember 2019 enden wird. Es werden neue Module des Digitalarchivs entwickelt (vor allem Zugriffmodul und mehrere Archivportalmodule), einige Module werden deutlich geändert und werden damit mehr leistungsfähig sein.

Parallel mit dem Projekt *National Digitalarchiv II* läuft ein Projekt des Innenministeriums, das sollte bisherige Zentralevidenz des Archivgutes und der Findmittel auf eine neue Ebene bringen und als ein Modul des Archivportals integrieren.

Bibliotheken und digitale Archivierung

Digitale Langzeitarchivierung war auch ein wichtiges Thema für tschechische Bibliothekare. Seit 2001 läuft ein Webarchiv in der National Bibliothek. Es entstand ein Projektteam in der National Bibliothek, der die Einrichtung der National Digitalbibliothek umsetzen sollte. Diese Fachleute haben viele Kenntnisse und Impulse aus dem Bereich der Digitalen Preservation aus dem Ausland nach Tschechien transferiert. Das Projekt der *National Digitalbibliothek* wurde von der National Bibliothek und der Mährischen Landesbibliothek in derselben Zeit wie das National Digitalarchiv realisiert, also in den Jahren 2011 bis 2015.

Ein nächstes Projekt in Kooperation von der Mährischen Landesbibliothek und der Masaryks Universität in Brünn war *LTP-Pilot* in den Jahren 2014 bis 2015, das auch Archivemata getestet hat.

Die Zusammenarbeit von der Nationalbibliothek, der Mährischen Landesbibliothek und der Bibliothek der Akademie der Wissenschaften ist auch heute die Triebkraft der digitalen Archivierung im Bibliotheksbereich. Hier muss ich noch das ARCLib-Projekt erwähnen, das zwischen Jahren 2015 und 2020 geplant ist und schon mehrere Ergebnisse bringt. Es sind zwei methodische Anleitungen/Richtlinien zur digitalen Langzeitarchivierung publiziert und es läuft die Entwick-

lung von speziellen Open Source Software für Bibliotheken, die grundsätzliche Funktionalitäten der Langzeitarchivierung anbieten sollte.

Die Konzeptionen der digitalen Archivierung in der Tschechischen Republik

Am 5. Dezember 2016 wurde die Nationalkonzeption der Langzeiterhaltung der digitalen Unterlagen in Bibliotheken publiziert, die im Zusammenhang mit der Konzeption der Entwicklung der Bibliotheken in Tschechien entstand. Dort wurden grundlegende Ziele genannt, wie zum Beispiel strategische Finanzplanung, Koordination zwischen Bibliotheken, Nutzung von ausländischen Lösungen usw.

Für unsere Zwecke ist sehr interessant eine ältere Version der Konzeption vom 6. November 2014. Dort gibt es noch scheinbare Anstrengung die Langzeitarchivierung nicht nur im Bereich von Bibliotheken, sondern in Kooperation mit anderen Ministerien zu koordinieren. Diese Version setzte Vorbereitung einer gemeinsamen Strategie für mehrere Ministerien und auch die Einrichtung eines gemeinsamen methodologischen Zentrums für digitale Archivierung voraus.

Auch das Innenministerium hat die Konzeption der Entwicklung von Archiven in der Tschechischen Republik von 2018 bis 2025 vorbereitet (aber nicht publiziert.) Man kann dort auch Ansätze für Langzeitarchivierung finden, aber es wird dort mehr der heutige Stand der Problematik beschrieben. Nach dem Archivgesetz gibt es nur das zentralisierte Digitalarchiv mit dem Archivportal im Nationalarchiv. Andere Archive (im Sinne des Archivgesetzes) können die Akkreditierung des eigenen Digitalarchivs anfordern.

Der heutige Stand der digitalen Archivierung

Heute gibt es in der Tschechischen Republik nur ein Digitalarchiv (laut dem Archivgesetz) und einige Gedächtnisinstitutionen haben Digital Repository im Sinne OAIS, unter anderen: Nationalbibliothek, Mährische Landesbibliothek, Bibliothek der Akademie der Wissenschaften, National Technische Bibliothek, National Filmarchiv, Archiv des Tschechischen Fernsehens, Archiv des Tschechischen Rundfunks, Archiv der Karlsuniversität, Archiv der Hauptstadt Prag und vielleicht noch andere.

Mehrere Gedächtnisinstitutionen (Regionalarchive sind auch dabei) haben digitale Speicher und bieten ihre Daten mittels eigenen Webapplikationen und Webportale an. (Archive typisch digitalisierte Archivalien und digitale Findmittel.) Es handelt sich nicht um Archivportale im Sinne des Archivgesetzes. In diesem Sinne gibt es nur das Nationalarchivportal im Nationalarchiv.

Es gibt Intentionen von einigen Institutionen und auch Personen eine intensive Zusammenarbeit und Erfahrungsvermittlung zu organisieren, aber diese bleiben oft fragmentiert. Hier kann ich zum Beispiel eine Aktivität vom Nationalarchiv erwähnen. Es handelt sich um eine Reihe von internationalen Workshops unter dem Namen Archives in Digital Age (AIDA), im Jahr 2018 unter dem Titel „Self-Audit and Certification of Digital Archives in Central European Perspectives“.

Es fehlt hier eine gemeinsame Plattform (wie zum Beispiel in Deutschland nestor), die alle Interessierenden um Langzeitarchivierung durch verschiedene Institutionen zusammenbringen und verschiedene Arbeitsgruppen für verschiedene Aspekte der digitalen Archivierung organisieren helfen könnte.

Die Perspektiven und Ziele der digitalen Archivierung im Archibereich

Es kommt langsam die Zeit, die Konzeption der digitalen Archivierung im Archibereich vorzubereiten. Diese Konzeption soll durch Archivlegislative und vor allem die Festlegung der langfristigen Finanzierung und Personalausstattung der digitalen Archivierung unterstützt werden.

Wir beginnen nicht von Anfang an. Wir haben schon vierjährige Erfahrungen mit dem Betrieb des National Digitalarchivs. Seit 2016 gibt es beim Nationalarchiv eine Methodische Plattform für elektronische Dokumente (Moped). Moped verbindet Archivare aus allen Regionalarchiven und aus einigen anderen Archiven, die versuchen, methodische und praktische Frage der digitalen Archivierung zu lösen. Wir haben viele gute Kontakte an ausländische Kollegen, wir nehmen regelmäßig an verschiedenen Internationalen Tagungen und Arbeitskreisen teil.

Im Forschungsbereich und bei der konkreten technologischen Lösungen und Werkzeugen sollte engere und intensivere Kooperation mit Universitäten und Akademie der Wissenschaft entstehen. Dasselbe gilt für Kooperation mit anderen Gedächtnisinstitutionen in der Tschechischen Republik. Es wäre vielleicht günstig die Idee einer gemeinsamen Zentralstelle für Koordination der digitalen Archivierung zu erfrischen.

Kompliziert bleibt die Frage der Zusammenarbeit mit dem Kommerzsektor, aber es scheint, dass es einen Weg durch die Angewandte Forschung zur praktischen digitalen Archivierung gibt.

Am Rande bleiben die Sicherheitsfragen, die oft Kompetenzen der digitalen Archivierung überschreiten. Hier werden Archivare auf die Lösungen vom IT Sektor und auf die methodische Hilfe der Sicherheitsexperten verlassen. Als eine mögliche Lösung (die jetzt nur im Nebel steht) wäre die Nutzung der vom Staat gesicherten IT-Infrastruktur (spezielle Netzwerke, Rechenzentren usw.) Dabei muss man aber an ganz spezifische Prinzipien der digitalen Archivierung und an daraus folgende Anforderungen für Speicherung und Zugriff zu digitalen Archivalien denken.

Digitale Archivierung in der Tschechischen Republik ist jetzt in der Gefangenschaft der DMS/VBS und der elektronischen Akten. Aus den Kapazitätsgründen bleiben am Rande wichtigere digitale Dokumente. Es ist nötig die Probleme von DMS/VBS zu verlassen und sich auf digitale Archivierung von Unterlagen, die nicht in die Analogform konvertierbar sind (Datenbanken, Vektorgrafik, GIS oder Web- und Intranetarchivierung) zu konzentrieren.

Zum Schluss möchte ich noch eine wichtige Bedingung, die für die Entwicklung der digitalen Archivierung nötig ist, hervorheben. Es fehlt die Verknüpfung an Universitätsstudiengänge. Langzeitig fehlen neue, innovative und qualifizierte Leute, die sich der digitalen Archivierung widmen wollen.

Summary

The presentation brings the brief history of the development of theoretical and practical solutions to the problem of digital archiving in the Czech Republic in the field of archives and libraries. The focus is on building the National Digital Archive as an infrastructure for public archives in the Czech Republic in the context of Czech legislation and e-Government. The current state of National Digital Archives and perspectives of its development are also mentioned. There are a lot of challenges in this area. It is time to develop the fundamental concept of digital archiving in Czech Republic, which will combine current status and tasks with good knowledge of international experience in the long-term preservation of digital documents.

DAS PROJEKT DIMAG. SACHSTAND 2019

Christian Keitel

Die Archivierungssoftware DIMAG wurde 2006 vom Landesarchiv zunächst für den eigenen Bedarf entwickelt.¹ Heute wird die Entwicklung für das DIMAG-Programmpaket von vier Partnern koordiniert und vorangetrieben und in über 70 Archiven eingesetzt. An dieser Stelle soll nun eine kleine Übersicht gegeben werden, was sich in der Zwischenzeit alles ereignet hat.

Beginnen wir mit den beteiligten Personen. 2006 wurde der erste DIMAG-Prototyp von zwei Archivaren und einem Informatiker erarbeitet. Heute wird die Entwicklung unter den Entwicklungspartnern abgestimmt. Das Landesarchiv Baden-Württemberg hat sich dafür mit dem Hessischen Landesarchiv, der Generaldirektion der Staatlichen Archive Bayerns und der Digitalen Archivierung Nord, kurz DAN, zusammengeschlossen. DAN wird von den staatlichen Archivverwaltungen in Hamburg, Bremen, Mecklenburg-Vorpommern, Sachsen-Anhalt und Schleswig-Holstein getragen. Bei jedem dieser vier Entwicklungspartner arbeiten Teams an DIMAG, die aus wenigstens einem Informatiker und einem Archivar bestehen. Beim Landesarchiv Baden-Württemberg besteht das DIMAG-Team derzeit aus vier Informatikern und drei Archivaren.

Eine Konstante in dieser Entwicklung ist der enge Austausch zwischen Archivaren und Informatikern. Gerade in einem noch immer so unbekanntem Gebiet wie der digitalen Archivierung ist es für die Weiterentwicklung entscheidend, dass die archivarischen Fachkenntnisse nicht den von der IT vorgegebenen Möglichkeiten und Begrenzungen widersprechen. Im Rückblick auf die letzten zwölf Jahre können wir feststellen, dass alle größeren Entwicklungsschritte letztlich auf einem ständigen Hin und Her zwischen den beiden Fachrichtungen beruhen, und dass dieses Hin und Her mit einem gemeinsamen tiefgehenden Verständnis der Fachaufgaben einherging. Die langjährige und unmittelbare Zusammenarbeit in den bereits genannten Teams war daher ein entscheidender Erfolgsfaktor. Aber auch alle anderen DIMAG einsetzenden Archive können

.....

1 Für die erste Zeit s. Christian Keitel, Rolf Lang und Kai Naumann, Konzeption und Aufbau eines Digitalen Archivs: von der Skizze zum Prototypen, in: Katharina Ernst (Hg.): Erfahrungen mit der Übernahme digitaler Daten. Elfte Tagung des AK „Archivierung von Unterlagen aus digitalen Systemen“ vom 20./21. März 2007, Stuttgart/Leipzig 2007, S. 36–41. Abrufbar auch unter http://www.staatsarchiv.sg.ch/home/auds/11/_jcr_content/Par/download-list_1/Down-loadListPar/download_14.ocFile/Text%20Keitel%20Lang%20Naumann%2Spei-che-rung.pdf [Abruf 29.5.2019]. Neuere Entwicklung s. auch Robert Kretzschmar und Christian Keitel, The Archiving System DIMAG and its Development through Cooperation and Collaboration: Policy, Professional and Technical Aspects, in: Karen Anderson, Irmgard Becker und Luciana Duranti (Hrsg.), *Born Digital in the Cloud: Challenges and Solutions. Contributions to the 21. Archival Science Colloquium /International Symposium of InterPARES Trust*, Marburg 2018, S. 83–97.

durch Ihre beständigen Rückmeldungen und Hinweise als Teilnehmer in dieser Entwicklung verstanden werden.

Eine Lernaufgabe stellte sich aber nicht nur innerhalb der Teams, sondern ebenso zwischen den einzelnen Teams. Die im Juli 2010 einsetzende Kooperation mit den Kolleginnen und Kollegen vom Hessischen Hauptstaatsarchiv Wiesbaden ging zwar zunächst von den in Baden-Württemberg entwickelten Konzepten aus. Bei der Weiterentwicklung musste nun aber jeder einzelne Schritt mit dem Team im benachbarten Bundesland abgestimmt werden. Rückblickend kann die Ausgestaltung der Kooperation auch als Lernprozess beschrieben werden. Zunächst ging es natürlich um das gegenseitige Kennenlernen, die Verständigung über die Punkte, die dem Gegenüber sehr wichtig waren und jene, die verhandelbar waren. Schritt für Schritt wurde aber auch eine ganze Reihe von Kommunikationskanälen eingeführt. Als erstes Kommunikationsmittel wurden auf der operativen Ebene Telefonkonferenzen vereinbart, die bis heute alle drei bis vier Wochen stattfinden. Daneben gab es gerade in der Anfangsphase verschiedene eintägige Treffen. Rückblickend gesehen war es sehr gut, dass wir zunächst nur mit einem Kooperationspartner begonnen haben. Wir hatten also schon unseren „Kooperationsgrundkurs“ hinter uns, bevor 2012 die Kollegen aus der bayerischen Archivverwaltung hinzukamen. Auf Leitungsebene wurde nun der DIMAG-Lenkungskreis eingerichtet, der sich seitdem zweimal im Jahr trifft. Die Leiter der vier Entwicklungspartner und die operativ verantwortlichen Mitarbeiter treffen dort die Grundsatzentscheidungen.

Mit der Einrichtung des Lenkungskreises erhielt auch die operative Ebene ihren Namen. In dieser Entwicklungsgruppe konzentrierten sich die Diskussionen schon bald auf das 2008 von Baden-Württemberg noch im Alleingang entworfene Metadatenmodell. Zunächst versuchten wir ab 2013 die anstehende Überarbeitung noch via E-Mail und eintägige Treffen zu bewältigen, ein Versuch, der nicht besonders erfolgreich war. Die Teams trafen sich daher am 26. und 27. Mai 2014 erstmals zweitägig zur „DIMAG-Metadatenklausur“, die intern bald in „Metadatenfolter“ umbenannt wurde. An diesen Treffen beteiligten sich von Anfang an auch die Kolleginnen und Kollegen vom DAN. Im April 2015 konnte schließlich die neue Metadatenreferenz verabschiedet werden. Das Gremium beschloss, diese frei zugänglich ins Netz zu stellen.² Aus der Metadatenfolter wurden nun die DIMAG-Entwicklungstreffen, die seitdem zweimal im Jahr jeweils zwei Tage lang stattfinden. Auf diesen Treffen werden alle wesentlichen Neuentwicklungen abgestimmt und erarbeitete Ergebnisse abschließend diskutiert.

In den ersten Jahren hat das Landesarchiv bereits die Archivierungskomponente von DIMAG entwickelt, die wir heute „DIMAG-Kernmodul“ nennen und außerdem die Übernahmesoftware IngestList. IngestList wird als einziges

.....
 2 Das Metadatenchema ist publiziert unter http://www.landearchiv-bw.de/sixcms/media.php/120/58735/DIMAG-Metadatenreferenz_2015-04-23.pdf [Abruf 29. 5. 2019]. Das Vorgängerschema findet sich unter https://www.landearchiv-bw.de/sixcms/media.php/120/48392/konzeption_metadaten10.28354.pdf [Abruf 29. 5. 2019]. Weitere Erläuterungen finden sich in Kai Naumann, Christian Keitel, Rolf Lang, A Metadata Concept for Mixed Digital Content at a State Archive, *The International Journal of Digital Curation* 2 Vol. 4 (2009), S. 80–92, <http://www.ijdc.net/index.php/ijdc/article/viewFile/120/123> [Abruf 29. 5. 2019].

DIMAG-Modul als Freeware abgegeben.³ Seit 2011 entwickelt die hessische Archivverwaltung in der Kooperation weitere Software für den Bereich des Ingests. Die bayerischen Kolleginnen und Kollegen übernahmen die Verantwortung für den Access-Bereich.

Als daher 2014 die Verhandlungen mit der Digitalen Archivierung Nord begannen, waren wesentliche Module bereits verteilt. Zugleich war der Koordinationsaufwand erheblich angestiegen. Vordringlich erschien daher die Aufgabe, gerade diesen Bereich personell zu verstärken. Der DAN-Beitrag besteht daher in der Finanzierung der Verfahrenspflegestelle, die organisatorisch beim Landesarchiv Baden-Württemberg in Stuttgart angedockt ist. Derzeit ist die Verfahrenspflegestelle mit zwei Informatikern besetzt. Es erscheint uns heute schlicht unvorstellbar, ein derartiges Projekt ohne eine derartige Stelle vorantreiben zu können.

Im Laufe der Jahre haben wir gelernt, wie Software gemeinsam und kooperativ entwickelt werden kann. Eine große Herausforderung ist dabei, dass jedes DIMAG-Team in eine eigenständige Archivverwaltung eingebettet ist. Jede Archivverwaltung kann auf eine lange Tradition zurückblicken. Da Archivalien in aller Regel Unikate sind, gab es auch sehr lange keine Notwendigkeit, sich abzustimmen. In dem DIMAG-Projekt mussten wir daher nicht nur die Grundregeln einer engeren Zusammenarbeit lernen. Zugleich hat jeder Entwicklungspartner eine eigene DIMAG-Instanz eingerichtet. Auch die praktischen Übernahmen und Nutzungsvorgänge laufen jeweils im Rahmen der einzelnen Archivverwaltungen ab. Diese Dichotomie – kooperative Softwareentwicklung bei eigenständigem Einsatz – hat für die Entwicklungspartner verschiedene Konsequenzen:

- Die Software muss offen sein für verschiedene technische Umgebungen. DIMAG läuft auf allen Linux-Derivaten.
- Die Software muss zu archivfachlich unterschiedlichen konzeptionellen Ansätzen passen.
- Die Entwicklungsteams erhalten aus ihren Archivverwaltungen und ebenso aus anderen DIMAG einsetzenden Archiven in der Regel sehr unmittelbare Rückmeldungen aus der Praxis, die sie dann wiederum in die Entwicklungsüberlegungen einfließen lassen können.

Bei allen in der digitalen Archivierung anstehenden Aufgaben muss ein Abgleich vorgenommen werden zwischen einer möglichen technischen Umsetzung (d.h. einer möglichen Automatisierung) und einer organisatorischen Umsetzung, die sich dann z. B. in Richtlinien ausdrücken kann. Besonders deutlich wird dies in der Frage, wie einzelne Metadatenfelder befüllt werden sollen. Ein Freitextfeld erlaubt allen Beteiligten, eigenen Vorstellungen nach Gutdünken nachzugehen. Zugleich ermöglicht diese Umsetzung auch die maximal denkbare Uneinheitlichkeit bei den eingetragenen Feldinhalten. Anders sieht es bei vom System ermittelten Feldinhalten aus, die keinerlei Abweichung durch das einzelne Archiv mehr zulassen. Eine Mittelposition nehmen Auswahlfelder ein, die nur eine begrenzte Zahl vorgegebener Inhalte zulassen. Mit anderen Worten legt ein Freitextfeld archivinterne Richtlinien sehr nahe, um wenigstens auf diesem Weg Einheitlichkeit zu erreichen. Bei einer systemseitigen Vergabe wurden dagegen die Richtlinien der unterschiedlichen Archivverwaltungen bereits vereinheitlicht und in das Programm eingebaut.

.....
³ Veröffentlichung: <https://www.landearchiv-bw.de/web/49289>. Software: <https://sourceforge.net/projects/ingestlist/> [Abruf 29. 5. 2019].

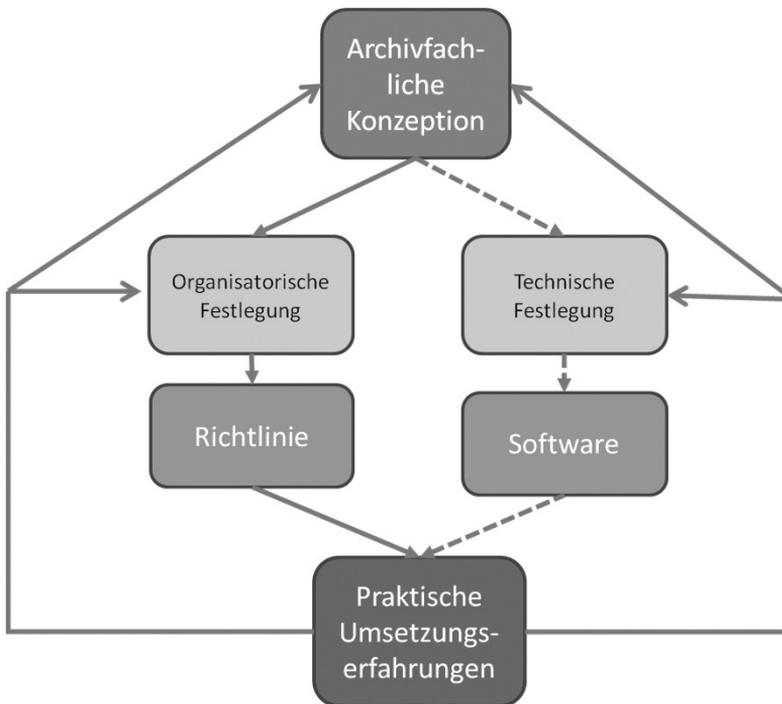


Abb. 1: Zusammenspiel von konzeptionellen Überlegungen, technischen Umsetzungen und praktischen Erfahrungen.

Auf der einen Seite profitieren die Entwicklungspartner also sehr von der unmittelbaren Nähe zur Praxis. Fehlerhafte Konzepte und Programmierungen fallen sehr schnell auf. Auf der anderen Seite entsteht immer wieder eine Spannung zwischen den Regelungen eines konkreten Archivs und den Absprachen im ganzen DIMAG-Verbund.

In den letzten Jahren wurden die Entwicklungspartner immer wieder von anderen staatlichen Archivverwaltungen gefragt, ob dieser Kreis nicht erweitert werden könne. Bereits die Einrichtung der Verfahrenspflegestelle deutete ja in die Richtung, dass der Koordinationsaufwand zwischen den entwickelnden Archivverwaltungen enorm gestiegen war. Der DIMAG-Lenkungskreis beschloss daher, die Zahl der Entwicklungspartner nicht erhöhen zu wollen. Auch DAN ist als Verbund nur ein Entwicklungspartner. Andererseits ermöglichte gerade der Verbund erst die Teilnahme der nördlichen Archivverwaltungen, denn DAN regelt die eigenen Belange intern. Anstelle von vier Entwicklungspartner sind es eben keine acht geworden.

Staatliche Archivverwaltungen waren aber nicht die einzigen, die an DIMAG interessiert waren. Unter den ersten Interessenten waren Kommunalarchive. Weitere Anfragen kamen von Kirchen- und Universitätsarchiven. Wie also sollten die Entwicklungspartner auf diese Anfragen antworten? Klar war, dass diese Häuser einen ganz anderen Unterstützungsbedarf haben als staatliche Archivverwaltungen. Zunächst überlegten wir uns bundesweite Lösungen, die dann auch mit einem für ganz Deutschland gültigen Gebührenmodell hätten arbeiten können. Diese Ideen wurden allerdings sehr schnell verworfen. Zu un-

terschiedlich sind die Archivstrukturen und bedürfnisse zwischen den einzelnen Bundesländern.

Stattdessen wurde vereinbart, dass jeder DIMAG-Entwicklungspartner für die an DIMAG interessierten öffentlich-rechtlichen und kirchlichen Archiven eigene Lösungen entwickeln kann. Um diese Archive von den bisherigen Partnern abgrenzen zu können, haben wir uns auf den Terminus der Anwendungspartner geeinigt. An andere Archive, zu denen auch die Archive der freien Wirtschaft gehören, wird DIMAG nicht abgegeben.

Für die Abgabe an Anwendungspartner haben die Entwicklungspartner eine Reihe von Grundsätzen aufgestellt:

- Mit an DIMAG interessierten Archive werden Verwaltungsvereinbarungen abgeschlossen.
- An Ausschreibungen beteiligen sich die Entwicklungspartner nicht.
- Die Software wird nur als Gesamtpaket an interessierte Archive gegeben.
- Jeder Entwicklungspartner legt in seinem Bundesland / seinen Bundesländern eigenständig die Konditionen fest. Er ist selbst zuständig für den Support „seiner“ Anwendungspartner und evtl. anfallende Gebühren.

Für die Form der Weitergabe wurden 2011 und 2013 drei Begriffe vorgeschlagen, die sich seitdem eingebürgert haben: Die Weitergabe der Software (Supportpartnerschaft), die Aufnahme des kleineren Archivs in die DIMAG-Instanz der staatlichen Archivverwaltung (Magazinpartnerschaft) und die Beauftragung eines Dienstleisters zum Betreiben der DIMAG-Instanz (Dienstleisterpartnerschaft).⁴

Die momentan größten Teilnehmerzahlen stammen aus Baden-Württemberg und Hessen. In Baden-Württemberg startete 2011 das Projekt „Kommunales DIMAG“.⁵ Die Einzelheiten wurden und werden in einer Projektgruppe besprochen, der Vertreter der Kreis- und Stadtarchive, der Rechenzentren und des Landesarchivs angehören. 2015 konnte dieses Vorhaben in die Realität umgesetzt werden. DIMAG wurde bei einem kommunalen Rechenzentrum installiert, das auch den technischen Support übernimmt. Der fachliche Support liegt beim Landesarchiv. Mit Stand 1. März 2019 archivieren in der beim kommunalen Rechenzentrum ITEOS installierten DIMAG-Instanz 17 Kreisarchive, 45 Stadtarchive und ein Gemeindearchiv selbständig ihre digitalen Archivalien. Jedes Archiv hat nur Einsicht und Zugriff auf die eigenen Archivalien. Jedes dieser Archive entrichtet an das Landesarchiv pro Jahr eine Gebühr in Höhe von 3000 Euro sowie einen weiteren Betrag an ITEOS für den Betrieb. Hinzukommen noch kleinere unselbständige Archive, die unmittelbar von den Kreisarchiven betreut

4 Christian Keitel, Das Rad zweimal erfinden? Kooperationsangebote des Landesarchivs Baden-Württemberg zur digitalen Archivierung, Digitale Registraturen – digitale Archivierung. Pragmatische Lösungen für kleinere und mittlere Archive? Beiträge zum 16. Archivwissenschaftlichen Kolloquium der Archivschule Marburg, Marburg 2012, S. 131–154. 31.

Christian Keitel, Dienstleisterpartnerschaft mit DIMAG, in: Kai Naumann und Peter Müller (Hrsg.), Das neue Handwerk. Digitales Arbeiten in kleineren und mittleren Archiven, Vorträge des 72. Südwestdeutschen Archivtags am 22. und 23. Juni 2013 in Bad Bergzabern, Stuttgart 2013, S. 54–57.

5 Zu dem Projekt s. Miriam Eberlein, Christian Keitel und Manfred Waßner, „DIMAG“ wird kommunal: Ein digitales Langzeitarchiv für Städte und Gemeinden in Baden-Württemberg, in: Digitale Archivierung. Innovationen-Strategien-Netzwerke, Mitteilungen des Österreichischen Staatsarchivs 59(2016), Wien, Innsbruck: Studienverlag 2016, S. 21–32.

werden. Seit 2017 erproben außerdem sieben der acht Universitätsarchive einen ähnlichen Betrieb. Hier wird die DIMAG-Instanz vom Rechenzentrum der Universität Tübingen betrieben. Auch die zwei evangelischen landeskirchlichen Archive und die beiden Diözesanarchive in Baden-Württemberg konnten sich auf das Kirchliche Rechenzentrum Südwestdeutschland als gemeinsamen Rechenzentrumsbetreiber einigen. Das Hessische Landesarchiv hat DIMAG dem Institut für Stadtgeschichte Frankfurt a. M. und dem Stadtarchiv Kassel sowie dem Digitalen Archiv der Hessischen Hochschulen gegeben, einen Verbund, in dem sich die Hochschularchive Frankfurt, Darmstadt, Gießen und Marburg zusammengeschlossen haben.

An all diesen Orten erproben die Kolleginnen und Kollegen nun DIMAG. Eine erste Folge war der erhöhte Supportbedarf. Für die baden-württembergischen Anwendungspartner richtete das Landesarchiv 2017 die Supportstelle ein. Etwa zur selben Zeit begannen die Entwicklungspartner das DIMAG-Wiki aufzubauen.⁶

- - Eine Dateisammlung rudimentär ordnen und erschließen
 - Verzeichnisstrukturen / Containerformate (z.B. ZIP)
 - DOC lässt sich nicht mehr mit Word öffnen
 - Unerwünschte Dateiformate ausfiltern mit IngestTool
 - Passende Anfängerprojekte
 - Löschen im Vier-Augen-Prinzip leicht gemacht
 - AID
 - Signatur
 - Alte Signatur nach Verschieben
 - Zuordnung Repräsentation – Informationsobjekt (IO)
 - Änderungen an der Struktur
 - Beschränkungen bei Dateiformaten im DIMAG
 - Löschen
 - Setzen des Status "Abgeschlossen"
 - Verkleinern großer PDF-Dateien
 - Bildung von AIPs (Archivierungsinformationsspaket)
 - Informationsobjekte
 - Feld Strukturtyp
 - Transport von IngestList nach DIMAG
 - Qualitätskontrolle nach Ingest
 - Integritätsprüfung (Prüfsummen etc.)
 - Unerwünschten Dateien in einem Ordnerbaum finden und löschen
 - Warum muss ich manche Dateien doppelt in DIMAG einstellen?
 - Aufbereitungsbericht
 - Best Practice – Benutzernamen in DIMAG
 - Massenhafte Konversion von Dateiformaten
 - Wie finde ich doppelte Repräsentationen?
 - Anlegen von Repräsentationen
 - Wie benenne ich Ordner massenhaft um?
 - Externe Metadaten weinternutzen

Eine Dateisammlung rudimentär ordnen und erschließen

Q: Ich habe eine große Menge Dateien bekommen, aber diese müssen korrekt sortiert und mit Titeln und eventuell Laufzeiten versehen werden.

Abb. 2: FAQ aus dem DIMAG-Wiki.

Derzeit kennt das Wiki den Kreis der Entwicklungspartner, der Betreiber (d. h. der Rechenzentren), der Anwendungspartner und die für die Öffentlichkeit gedachten Informationen. Bis auf den zuletzt genannten Bereich sind alle Bereiche passwortgeschützt. Im Bereich der Entwicklungspartner stehen für den Zeitraum März 2013 bis Ende Februar 2019 die Protokolle von 18 Treffen und

6 <https://dimag-wiki.la-bw.de/xwiki/bin/Public/> [Abruf 29. 5. 2019].

insgesamt 108 Telefonkonferenzen. In sechs Jahren haben wir durchschnittlich also alle drei Wochen miteinander telefoniert. Der Bereich der Anwendungspartner ist (ebenso wie die anderen Bereiche) nicht weiter nach den verschiedenen soeben genannten Sparten oder Bundesländern geteilt. Schließlich gehen wir davon aus, dass diese Gruppen wechselseitig von ihren gemachten Erfahrungen profitieren können. Umso mehr gilt dies für das DIMAG-Forum, das 2018 eingerichtet werden konnte. Während das Wiki eher statische Informationen (z. B. Protokolle, Handbücher) enthält, werden hier offene Fragen kontrovers diskutiert und nicht zuletzt auch beantwortet. Das Forum und auch das Wiki haben schnell an Umfang gewonnen. Zugleich haben wir in der DIMAG-Supportstelle den Eindruck, dass der unmittelbare Supportaufwand für einzelne Archive gesunken ist, da sich vieles schon in diesen neuen Medien nachlesen lässt.

Es ist kein Zufall, dass sich die meisten Anwendungspartner in Verbänden organisiert haben. Als verhältnismäßig kleine Archive haben sie praktisch keine Möglichkeit, selbständig in die digitale Archivierung einzusteigen. Und wenn sie doch die Software selbst erwerben und betreiben bzw. betreiben lassen können, stellen sich sehr viele Fragen, wie diese Software in der Praxis eingesetzt werden soll. Erneut muss das Zusammenspiel von Software und hausinternen Regelungen sowie Richtlinien justiert und ausgerichtet werden. Hier kann ein Verbund mit eigenständigen Rückversicherungsmöglichkeiten enorm viel helfen.

Und dann gibt es noch die Anfragen von Archiven außerhalb der Bundesländer, deren Archivverwaltungen selbst DIMAG einsetzen. Auch hier ist es den Entwicklungspartnern nicht möglich, jedes einzelne Archiv umfassend selbst zu betreuen. Stattdessen wird der Aufbau von Verbänden angestrebt. Gerade in diesen Fällen ist es besonders wichtig, dass sie mit den anderen Beteiligten, zu denen wir auch die staatlichen Archivverwaltungen aus diesen Bundesländern zählen, abgesprochen werden.

Vergleichbare Anfragen haben uns auch aus Österreich und der Schweiz erreicht. Die DIMAG-Entwicklungspartner haben um die Bildung von Verbänden gebeten. Voraussichtlich können die beiden Verträge noch 2019 unterzeichnet werden. Die Entwicklungspartner haben dann in Österreich und der Schweiz jeweils einen Ansprechpartner. In dem Gesamtgefüge der DIMAG-Entwicklung werden die beiden Verbände als Aufgabe die Durchsicht der Software auf mögliche Fehler und Inkonsistenzen übernehmen (sog. sourcecode review). Im Zuge dieser Verhandlungen haben sich die Entwicklungspartner auch darauf festgelegt, DIMAG weiterhin als deutschsprachige Software betreiben zu wollen.

Mit der steigenden Zahl der Anwendungspartner stellt sich immer dringlicher die Frage, wie die Entwicklungspartner auf Wünsche um Individualentwicklungen reagieren sollten. Ausgangspunkt ist der Grundsatz, dass die DIMAG-Software einheitlich und stringent weiterentwickelt werden sollte. Über diese Fragen entscheiden die vier Entwicklungspartner. Sie übernehmen auch einen Großteil der anstehenden Aufgaben. Zugleich sind die DIMAG-Entwicklungspartner in einem ständigen Austausch mit allen anderen DIMAG einsetzenden Archiven. Auf diesem Weg sind schon viele Anregungen in die DIMAG-Entwicklung eingeflossen. Dabei konnten die Entwicklungspartner überzeugt werden, dass das Anliegen im Interesse aller DIMAG einsetzenden Archive liegt. Gleichwohl sind die DIMAG-Entwicklungspartner nicht verpflichtet, die Wünsche umzusetzen. Dieses Überzeugen fällt leichter, wenn die Wünsche nicht von einem einzelnen Archiv, sondern von einem Verbund vorgetragen werden. Nicht selten ist es aber

auch so, dass die Anliegen mit der bereits bestehenden Software umgesetzt werden können. Auf jeden Fall gewinnen alle Parteien von dem umfassenden Dialog, der durch ganz praktische Anwendungsfälle ausgelöst wird. Diese Entwicklung bestätigt eine These, die seit jeher eine wesentliche Grundlage von allen Kooperationsverträgen war: Diese Verträge beschreiben nicht die Abgabe von Software, sondern eine darüber hinausgehende viel umfassendere Kooperation in allen Angelegenheiten rings um die digitale Archivierung bis hin zur letztlich gemeinsamen Entwicklung der Software.

Konzeptionelle Entwicklung

Was hat sich nun seit 2006 bei der Software selbst getan? Von Beginn an sollte die Entwicklung eines digitalen Archivs als „mehrstufiges Prototyping“ vorgenommen werden. Nachdem das heutige DIMAG 3.0 immer noch weiterentwickelt wird, kann man diese Annahme wohl als bestätigt ansehen. Allerdings sehen wir nun eher einen kontinuierlichen Entwicklungsprozess. Frühere Hoffnungen, dass die Entwicklung nach wenigen Jahren weitgehend abgeschlossen werden könne, haben sich nicht erfüllt.

Gleichwohl haben sich einige Themen geändert:

- 2006 und 2007 unterhielten wir uns noch sehr intensiv über geeignete Betriebssysteme und Datenträger. Wir entschieden uns dann für Festplattensysteme. Beides ist immer noch ein Thema für uns, jedoch in etwas veränderter Form, seitdem die meisten DIMAG-Installationen bei Rechenzentren installiert wurden.
- Anfang hatten wir noch sprechende Dateinamen angestrebt, um für den Fall der Fälle, also für Zuordnungsprobleme, gerüstet zu sein. Mit DIMAG 2.0 stiegen wir dann 2009 konsequent auf nichtsprechende Dateinamen um, die dafür einen dauerhaften Identifier enthielten. Er kann nicht nur die Kommunikation zwischen verschiedenen IT-Systemen innerhalb des Archivs beitragen. Darüber hinaus bietet er auch eine Möglichkeit, einen im Internet eindeutigen persistenten Identifier wie beispielsweise URN oder DOI zu bilden.
- Während 2007 die Eingabe der übernommenen digitalen Archivalien noch hauptsächlich über ein Eingabeformular erfolgte, werden nun die seitdem entwickelte Upload-Möglichkeiten eingesetzt, sobald mehrere oder viele Archivalien auf einmal hochgeladen werden sollen.

Vielleicht die längsten Diskussionen gab es über die Abbildung der Tektonik in DIMAG. Vor gut zehn Jahren unterhielten wir uns noch darüber, ob es eine gemeinsame Tektonik für konventionelle und digitale Archivalien geben sollte und ob Repräsentationen wenigstens bei letzteren ein sinnvolles Abbildungsmittel darstellen könnten. Beide Diskussionen werden seit längerem nicht mehr in unserer Community geführt. Gefragt wird allerdings, ob die Tektonik wirklich auch in DIMAG abgebildet werden soll, wie wir das ja auch seit 2006 tun. Gegen dieses Konzept sprechen zunächst die Aufwände, Metadaten an zwei Orten nachzuweisen und diese zu synchronisieren. Dennoch finden wir auch nach den Diskussionen der letzten Jahre nach wie vor einige Argumente, die uns auf unserem bisherigen Weg bestätigen.

Unterlagen werden zumeist sehr kreativ erstellt, d. h. die folgen, wenn überhaupt, nur wenigen Regeln. Während die einzelnen Objekte sich oft nicht

eindeutig voneinander abgrenzen lassen (beispielsweise Dateisammlungen), beziehen sie sich schon über ihre Struktur aufeinander. Die Akten einer abgehenden Stelle lassen sich am besten im Gesamtzusammenhang dieser Stelle und innerhalb der vom Aktenplan vorgegebenen Strukturen abbilden. Es geht also nicht ausschließlich um das einzelne Objekt, wie dies zumeist bei Bibliotheken der Fall ist, sondern auch und ganz wesentlich um den Gesamtzusammenhang der Objekte. Dieser Zusammenhang kann gut monohierarchisch abgebildet werden, also graphisch durch einen Baum. Diese Struktur wird wiederum durch den einschlägigen Erschließungsstandard ISAD(G) unterstützt. Im DIMAG erfüllt die Tektonik damit folgende Zwecke:

- Keine willkürliche Trennung der einzelnen Bestandteile eines digitalen Archivals (z. B. Akte – Vorgang – Dokument).
- Orientierung der Archivare in ihrer gewohnten Umgebung.
- DIMAG soll weiterhin auch als alleinstehende System, („standalone“) verwendet werden können. Auch in diesem Fall muss eine Orientierung in DIMAG möglich sein.
- Es ist wahrscheinlich, dass die in den nächsten Jahren aufzubauenden Funktionalitäten zur digitalen Bestandserhaltung nicht im jeweiligen AFIS sondern in den Archivierungssystemen umgesetzt werden. Die anstehenden Erhaltungsentscheidungen können leichter vor dem Hintergrund der Tektonik getroffen werden.
- Das Archivierungssystem enthält alle Bestandteile des Archivierungspakets (AIP), also alles, was zum Verständnis und zur erneuten Darstellung des Archivals notwendig ist. Die Archivarinnen und Archivare sind daher in Erhaltungsfragen nur von einem System abhängig.
- Die einzelnen Bestandteile des Archivierungspakets unterliegen denselben Maßnahmen, z. B. derselben Art an Protokollierung. In den heute gebräuchlichen Archivischen Fachinformationssystemen gibt es keine vergleichbare Protokollierung.

Rückblickend gesehen kann die Entwicklung von DIMAG vielleicht in bislang drei Phasen unterteilt werden:

Zunächst gab es genau ein Programm, das sich DIMAG nannte und in dem die digitalen Archivalien abgelegt und archiviert werden konnten. Dieses Programm stellte 2006 die Handlungsfähigkeit des Landesarchivs Baden-Württemberg bei der Übernahme und Archivierung sicher. Allerdings bildete es nicht alle Aufgaben der digitalen Archivierung ab. Gerade das Funktionsmodell von OAS führt in seiner detaillierteren Form zahlreiche mit der digitalen Archivierung zusammenhängenden Aufgaben vor Augen.

Bereits 2007 begann das Landesarchiv Baden-Württemberg daher, DIMAG die bereits erwähnte Software IngestList an die Seite zu stellen, zu der nach Abschluss der Kooperationsverträge mit den Kolleginnen und Kollegen aus Hessen und Bayern noch das IngestTool und das AccessTool stießen. Vier einzelne Programme also, die zugleich auf verschiedene Weise miteinander verbunden sind. DIMAG bezeichnete nun das Gesamtpaket.

Seit einigen Jahren stellen wir zunehmend fest, dass immer mehr Funktionalitäten dieser Module an verschiedenen Stellen benötigt werden. Dass diese Teile also flexibel und unabhängig von den restlichen Programmteilen ihrer bisherigen Programmumgebung benutzbar sein sollten. Zugleich benötigen wir mit

dem Kernmodul nach wie vor einen stabilen Anker für alle diese Anforderungen, auch haben wir einheitliche Vorgaben für die verschiedenen sich darum gruppierenden Prozesse formuliert. Das Gesamtpaket beginnt sich daher in drei Bereich aufzugliedern:

In der Mitte das Kernmodul (also die Archivierungseinheit im engeren Sinne).

Umgeben wird es von Workflow-Engines, in denen wiederum kleinere und flexiblere Tools eingehängt werden.⁷ Ganz ähnlich sieht auch die Architektur vor, die die großen internationalen Hersteller Archivemata und Preservica in einem gemeinsame Memorandum of Understanding entwickelt haben. Beide Hersteller gehen davon aus, dass die Tools zunehmend systemübergreifend und oft auch als Open Source entwickelt werden.⁸

Jenseits dieser großen Trends gehen die Arbeiten „im Kleinen“ weiter. Beim Staatsarchiv Hamburg wird das Bestandserhaltungsmodul BEM entwickelt und wahrscheinlich zur Jahresmitte auch in einer ersten Version fertiggestellt. Es ermöglicht die kontrollierte und protokollierte Migration von Dateien. Beim Landesarchiv Baden-Württemberg ging 2018 die seit 2006 eingespielte BOA-Kooperation zu Ende.⁹ Das Landesarchiv hat daher das Webarchivierungstool DiWI entwickelt, das ebenso wie das BEM nach einer Erprobungszeit an die anderen DIMAG-Partner abgegeben werden soll.

Zusammenfassend können wir vielleicht bemerken, dass das Projekt „DIMAG“ auf einem guten Weg zu sein scheint. DIMAG ist ein Software-Projekt. Ausschlaggebend waren und sind aber die Menschen, die ein solches Projekt stemmen. An dieser Stelle sei daher allen Partner ganz herzlich für ihr Wohlwollen und ihre Unterstützung gedankt. Es ist schön, immer noch sagen zu können, dass wir gemeinsam nachweislich mehr erreichen können als jeder für sich alleine.

Summary

DIMAG is a software suite for the long-term preservation of digital records. The 1st version has been developed in 2006 by the Landesarchiv Baden-Württemberg (regional state archive). Nowadays, eight state archives share the burden of the further development of the software.

The software comprises different modules containing the main OAIS functional entities like ingest, data management, archival storage, preservation planning and access. The article discusses the mutual relationship between the conceptual assumptions of the archivists and the information scientists as well as the interplay between the development teams and the concrete users of the software. The AIP concept, for example, has been aligned to the hierarchical concepts of ISAD(G). During the years some decisions in the software design have been confirmed, others have been revised.

7 Vgl. den Beitrag von Frank Obermeit in diesem Band.

8 Matthew Addis, Justin Simpson, Jonathan Tilbury, Jack O’Sullivan, Paul Stokes, Digital Preservation Interoperability through Preservation Actions Registries, <http://doi.org/cq9p> [Abruf 29. 5. 2019].

9 Zu BOA s. Felix Geisler, Wiebke Dannehl, Christian Keitel und Stefan Wolf, Zum Stand der Webarchivierung in Baden-Württemberg, *Bibliotheksdienst* 51 (2017) H. 6, S. 481–489

DIMAG is a community project. Some remarks are given on the evolution of this community. The colleagues discuss the policies, the development and the deployment of the software in different circles and within the DIMAG wiki and the DIMAG forum. The article describes the development of these forms.

BEWERTUNG
UND
ÜBERNAHME

QUALITATIVE BEWERTUNG IM HINBLICK AUF INFORMATIONSGEHALT, VERWENDUNG UND KOSTENMETRIK¹ JUNI 2019

Martin Kaiser

Bewerten, die archivische Königsdisziplin

Die archivische Bewertung gilt als Königsdisziplin innerhalb der Abläufe im Archiv. Ohne Bewertung käme es zu einer unüberblickbaren Informationsflut mit erheblichen Redundanzen oder zu unkontrollierten Überlieferungsverlusten.

Interessanterweise ist die Bewertung in der analogen Welt eine binäre Entscheidung: Archivieren oder Kassieren. Um diese zwei Möglichkeiten haben sich eine Reihe von Methoden und Theorien entwickelt, und bisherige Überlegungen zur Bewertung haben sich auf die Unterstützung des Bewertungsprozesses konzentriert, mit dem Ziel der Objektivierung, Automatisierung, Beschleunigung und Vereinheitlichung.² Zwischen Nehmen und Nicht-Nehmen gab oder gibt es bisher aber keine Alternative.

Exkurs in die Natur digitaler Daten

Bei der Aufzeichnung analoger Informationen, Bild, Ton oder Bewegtbild, wird immer ein analoger Informationsstrom in einen diskontinuierlichen digitalen Datenstrom umgewandelt. Das analoge Signal (eine Farbe, ein Ton, eine Bewegung) wird periodisch abgetastet und digitalisiert; die Frequenz des Ab tastens und die Qualität der Analog-Digital-Wandlung bestimmen die Qualität des Resultats, aber auch die Grösse des entstehenden digitalen Datenstroms.

-
- 1 Schriftliche Fassung des Referats an der 23. Jahrestagung des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen» in Prag, 12./13. März 2019. Ich danke den Tagungsteilnehmenden für die angeregte Diskussion. Mein Dank geht ferner an meine Kollegin Claire Röthlisberger und meinen Kollegen Georg Büchler für die gemeinsamen vorbereitenden Überlegungen und für ihre Korrektur der schriftlichen Fassung.
 - 2 Sehr ausführlich beschäftigt sich der „Arbeitskreis Archivische Bewertung“ im VdA – Verband deutscher Archivarinnen und Archivare mit dem Thema: <https://www.vda.archiv.net/arbeitskreise/archivische-bewertung.html> [Abruf 4. 10. 2019].

Der archivistische Umgang mit Digitalisaten im Bereich Retrodigitalisierung und *Digital Born Data* ist zunächst unter diesem Gesichtspunkt zu betrachten, bevor wir wieder zum Thema Bewertung zurückkehren.

Maximale Qualität

Lange Zeit galt unwidersprochen die Übernahme und Archivierung in höchstmöglicher und -verfügbarer Qualität als Ziel der digitalen Archivierung bzw. Retrodigitalisierung.

Aber gerade im Bereich der Retrodigitalisierung hat die Entwicklung der Scannertechnologie und der Kameras³ in den letzten Jahren dazu geführt, dass die heute höchstmögliche Qualität je nach Vorlage und technischer Ausrüstung weit über die sinnvolle Qualität hinausgehen kann. Weil die Auflösung und Farbtiefe des Materials in direktem Zusammenhang mit den anfallenden Speicherkosten stehen, ist bei vielen Archiven inzwischen auch ein Bewusstsein für die Kostenmetrik angekommen. Das Gleiche beobachten wir im Bereich der Audio-, Video- und Film-Digitalisierung.

Datenspeicherung und Kostenmetrik

Die Datenspeicherung, im OAIS-Modell mit *Archival Storage* bezeichnet, wird in der Informatik meist mit dem enger gefassten Begriff *Bitstream Preservation* umschrieben. Die technischen Voraussetzungen und Herangehensweisen haben sich in diesem Bereich weitgehend vereinheitlicht und sind damit auf eine Weise quasi-standardisiert worden, dass wir heute von Speicherplatz von einem zu einem bestimmten Preis verfügbaren Rohstoff sprechen können.

Damit ist die Datenspeicherung ein Skalengeschäft geworden, der Preis sollte mit zunehmender Speichermenge abnehmen. Auch würde *Kryder's Law*⁴ einen fortlaufenden Fall der Speicherkosten erwarten lassen.

Die reale Erfahrung im Umfeld der Archive zeigt aber, dass der Preiszerfall nicht unbegrenzt weiterzugehen scheint. Mögliche Erklärungen sind, dass die Basiskosten (*Housing, Energy, Administration, Security*) kontinuierlich steigen oder dass aus spezifischen technischen Gründen ein Engpass bei der Datenübertragung entstanden ist.⁵

Digitale Fotografie / Retrodigitalisierung

Bei digitaler Fotografie gibt es eine ursprüngliche maximale Qualität, die von den Eigenschaften des Aufnahmesystems und der integrierten Nachbar-

.....
3 Auflösung oder Rasterung, Qualität der analog-digital Wandlung und Geschwindigkeit dieser Wandlung haben kontinuierlich zugenommen.

4 *Kryder's Law* meint die Beobachtung des schnellen Wachstums der Speicherdichte von magnetischen Speichermedien; siehe https://en.wikipedia.org/wiki/Mark_Kryder [Abruf 4. 10. 2019].

5 So ist der grösste Kostenblock bei der Datenspeicherung in Amazon Glacier die Kosten für Up- und Download der Daten.

arbeitsprozesse abhängt. Diese vorgegebene Qualität kann aber im Hinblick auf eine mögliche spätere Nutzung und im Hinblick auf Kostenersparnis durchaus verringert werden. Bei vielen Kameras im Consumerbereich ist diese Funktion bereits in der Kamera integriert, exportiert wird eine kleiner gerechnete JPEG-Datei.

Bei der Retrodigitalisierung gilt es heute, die optimale Qualitätsstufe für das digitalisierte Original zu finden. Daneben stellt sich auch die Frage nach einer geeigneten, allenfalls auch verlustbehafteten Komprimierung, gerade im Hinblick auf den Verwendungszweck.

Beispiel: Retrodigitalisierung

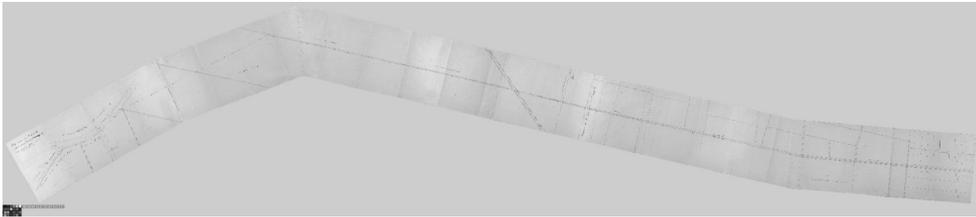


Abb. 1: Plan einer Telefonleitung: TIFF-Datei 2.13 GB⁶

Das Bild ist ein Scan eines Plans für den Bau einer Telefonleitung zwischen Biel und Bern. In der Vergrösserung sehen wir die Telefonleitung rot, blau koloriert ein Kanal und daneben ein Fahrweg. Der Plan ist zusammengeklebt, darum ist auf zwei Drittel des Bildes nur der Scantisch zu sehen.

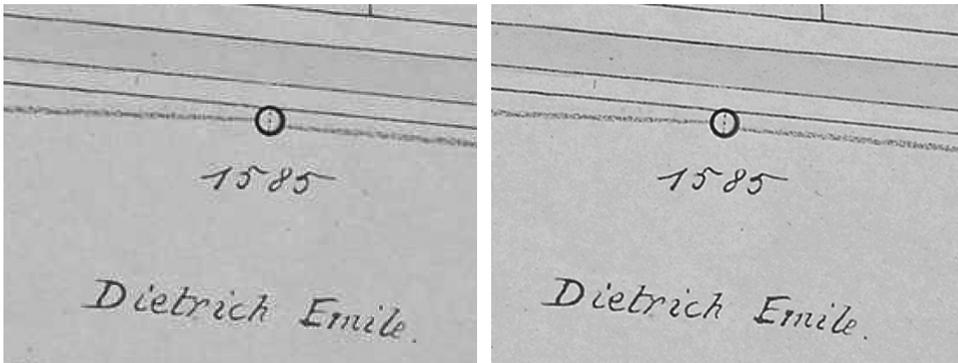


Abb. 2: Komprimiert JPEG2000 – 89.14 MB

JPEG – 2.44 MB

Die geringe Informationsdichte (nur Striche und einheitlich gefärbte Flächen, zudem der grosse Anteil des Scantischs am Bild) ruft nach einer Datenkomprimierung. Eine Datenkomprimierung mit LZW in TIFF erreicht aber nur etwas mehr als Faktor 2. Der Grund dafür ist, dass die eigentlich einheitlich gefärbten Flächen durch Rauchen im Aufnahmesensor aus vielen leicht unterschiedlich gefärbten Punkten bestehen. Erst verlustbehaftete Komprimierung bringt den gewünschten Effekt. Die 20-fache JPEG2000-Komprimierung ist

.....
6 Der Scan wurde uns freundlicherweise vom Staatsarchiv Bern zur Verfügung gestellt.

visuell verlustfrei⁷. Die JPEG-Komprimierung um den Faktor 100 zeigt Artefakte⁸, ist also nicht visuell verlustfrei, von einem Informationsverlust kann jedoch nicht die Rede sein.

Digital Born

Auch bei genuin digitalen Daten gibt es das Phänomen sehr unterschiedlicher Datenmengen für die gleiche Information. Am eklatantesten ist dies im Bereich vektorbasierter Daten wie CAD (*Computer-Aided Design*), GIS (Geoinformationssysteme) und Vektorgrafiken, wo der Unterschied zwischen der vektorbasierten Speicherung (z. B. in PDF/A) und einer Bitmap-Speicherung gut den Faktor 20 erreichen kann.

Bei Daten im Audio-, audiovisuellen und Bildbereich ist zu berücksichtigen, ob die Daten bereits an der Quelle in einem komprimierten Format entstanden sind (JPEG, MP3 etc.) – eine Konvertierung in ein archivtaugliches Format (TIFF, WAVE) würde in diesem Fall nur zu einem Anschwellen der Dateigröße führen.

Beispiele: Digital Born Data



Abb. 3: Ein privates Foto von der 17. AUdS-Tagung am 13./14. März 2013 in Dresden veranschaulicht das Thema. Das von der Kamera exportierte Bild ist eine JPEG-Datei von ungefähr 0.5 MB. Vor dem Datenexport wurde das Foto wahrscheinlich als RAW-

-
- 7 *Visually Lossless*, d.h. ein menschlicher Betrachter kann auch bei genauer Kontrolle in der Originalgröße keinen Unterschied erkennen (in der Abbildung sehen wir eine zweifache Vergrößerung).
 - 8 JPEG ist für Striche und Flächen nicht der optimale Komprimierungsalgorithmus, es entsteht entlang der Kanten ein sogenannter *Moiré-Effekt*.

Datei intern gespeichert. Ein Zugriff auf die oder ein Export der RAW-Datei ist nicht vorgesehen, aber möglicherweise wäre ein Datenexport auch als TIFF-Datei denkbar. Eine nachträgliche Umwandlung in eine unkomprimierte TIFF-Datei erzeugt eine Datei von 15 MB, mit LZW-Komprimierung kommen wir auf etwas mehr als 6 MB.



Abb. 4: Das zweite Beispiel zeigt einen Export aus einem GIS System⁹: eine Gewässerkarte des Kantons Zürich, freundlicherweise zur Verfügung gestellt von der GIS-Fachstelle des Kantons Zürich. Wir betrachten einen jeweils gleichen kleinen Ausschnitt um den Ort Fischenthal, damit die Unterschiede erkennbar werden. Ein Export als Vektor-PDF ist 1 MB gross und erlaubt ein beliebiges Skalieren der Karte.



Abb. 5: Das zweite Beispiel rechts zeigt den gleichen Ausschnitt als Bitmap-PDF. Die Karte ist hier als JPEG Bild eingebettet und – obwohl 14 MB gross – eher von bescheidener Qualität. Nicht abgebildet ist der Export als TIFF-Datei. Diese wird 66 MB gross und unterscheidet sich qualitativ nicht gross vom JPEG im PDF.

.....
9 In diesem Fall ArcGIS von der Firma ESRI

Bewerten im digitalen Umfeld

Was lässt sich aus den oben angestellten Betrachtungen schliessen?

Bewerten bedeutet im digitalen Umfeld nicht nur Übernehmen oder Kassieren, sondern wird zu einer Entscheidung für das für den Archivierungszweck angemessene Format und die dem Informationsgehalt angemessene Datenmenge. Bei „angemessener Datenmenge“ sprechen wir gemeinhin auch von Qualität.

Hohe Qualität bedeutet in der Regel höhere Datenmenge, das heisst höhere Kosten oder bei gleichen Kosten weniger archivieren. Die Beispiele zeigen freilich, dass ein unglücklich gewähltes Format auch mit hoher Datenmenge einhergehen kann.

Die Bewertung von digitalen Daten muss demnach nicht nur den rechtssichernden Charakter, die Nachvollziehbarkeit des Verwaltungshandelns und die Interessen der historischen Forschung berücksichtigen, sondern auch Qualität, Formatwahl und längerfristige Speicherkosten einbeziehen und auch diese Entscheidung in entsprechenden Bewertungsentscheidungen transparent machen.

Summary

Is there any sense in archiving nonsense, respectively signal noise?

Appraisal, especially of audiovisual data, does not only focus on the type of the accession (integral, selection, sampling or scrapping) but is getting more and more a question of amount of information acquired per archiving object. For a long time, the undisputed goal was acquisition and archiving in the highest quality possibly available. The development of scanner technology in recent years, however, has shown that the highest possible quality today, depending on the original and the technical equipment, can go far beyond any meaningful quality. Because the amount of information or quality of the material is directly linked to the accumulated storage costs, many archives do realize that „ten times less storage space costs ten times less“.

There are some interesting topics in this context:

For digitally born audiovisual material, there is an original maximum data quality that depends on the characteristics of the recording system. Any post-processing usually reduces this quality to get a final product. This material of reduced quality may well be used for archival purposes in respect for long-term cost savings.

In the case of *retro-digitization* on the other hand, it is important to find the right quality level (resolution and color depth) with respect to the digitized original, so that no unwanted signal noise is generated by the recording system and filling the files.

In addition, there is also the question of a suitable, possibly also lossy, compression, especially with regard to the intended later use.

In essence: Appraisal in the digital world means that archives must also provide tools and guidelines for qualitative evaluation to keep longer-term storage costs under control.

ÜBERNAHME VON LANDWIRTSCHAFTLICHEN DATEN AUS FACHAPPLIKATIONEN

... und etwas Geschichte der schweizerischen Agrarpolitik

Flurina Camenisch, Bernhard Stüssi

Agrarpolitik in der Schweiz und Einführung der Direktzahlungen

Die Kantone der Schweiz und ihre Rechtsvorgänger erheben seit ihrem Bestehen eine Vielzahl von statistischen Daten zu Kontroll-, Beitragszahlungs-, Besteuerungs- und Planungszwecken. Mit der Gründung des modernen Bundesstaates 1848 setzte unter der Ägide des Eidgenössischen Departements des Innern auch eine schweizweite Systematisierung und Vereinheitlichung von Datenerhebung und Statistik ein.

Seit ihrem Beginn ist auch die Agrarpolitik auf solche Erhebungen angewiesen. Im Staatsarchiv Graubünden finden sich kantonale Viehzählungen seit 1835, worin zunächst nur der „Viehbestand der einzelnen Gemeinden“ mit jeweils einer Zahl angegeben ist.¹ Ab den 1860er-Jahren erhoben die kantonalen Behörden im Auftrag des Eidgenössischen Statistischen Büros wesentlich detailliertere Zahlen. Bei den Viehzählungen wurden die Kategorien Pferde, Rindvieh und Schweine sowie pauschal Schafe, Ziegen und Bienenvölker erfasst. In Zusammenarbeit mit dem Bund sammelte auch das als privater Verband organisierte Schweizerische Bauernsekretariat eine Fülle landwirtschaftlicher Strukturdaten, die es unter dem Titel „Landwirtschaftliche Monatszahlen“ publizierte. Darin finden sich unter anderem Angaben zu Betriebsgrössen, Viehbestand, Produktionsmengen, Preisverhältnissen und Witterungseinflüssen auf regionaler, kantonaler und Landes-Ebene.² Im Rahmen der Betriebszählungen wurden ab 1907 die Landwirtschaftsbetriebe mit der Anzahl ihrer Beschäftigten, Besitzverhältnissen, Angaben zur Bodenbenützung und zum Viehbestand sowie zu den vorhandenen technischen Hilfsmitteln separat erfasst.³

Die während und nach dem Ersten Weltkrieg aufgetretenen Versorgungsengpässe des Binnenlands Schweiz führten zu einem Ausbau der Agrarpolitik.

1 StAGR X 6 s.

2 StAGR X 3 e.

3 StAGR II 12 L.

Die Landwirtschaft wurde zunehmend als „Service public“ im Ernährungsbereich, als Gewährleisterin der Ernährungssicherheit, gesehen. Insbesondere mit der im Zweiten Weltkrieg durch die Bundesbehörden organisierten so genannten „Anbauschlacht“ flossen vermehrt planwirtschaftliche Elemente in die Agrarpolitik ein. Kein Zufall ist es daher, dass ab 1939 im Staatsarchiv Graubünden unter der Signatur StAGR X 3 x Unterlagen zum „Landwirtschaftlichen Produktionskataster“ vorhanden sind. Darin werden unter anderem Grösse und Beschaffenheit jeder „Flurabteilung“ (bewirtschaftete Fläche nach Örtlichkeit) detailliert beschrieben.⁴

Nach dem Zweiten Weltkrieg stand nebst der Ernährungssicherheit vermehrt der Schutz der heimischen Landwirtschaft vor Markteinflüssen, insbesondere vor günstigen Importen, im Vordergrund. Zu einer zentralen Masseneinheit dieser Politik wurde der so genannte „Paritätslohn“: Bauern- sollten denselben Lohn erhalten wie Industriearbeiter-Familien.⁵ Mit der 1954 in Kraft getretenen Landwirtschaftsverordnung wurden als wichtigstes Mittel zur Erreichung dieses Ziels Preisstützungen, zunächst auf Getreide, Milch und Zuckerrüben, eingeführt.⁶

Der Markteingriff des Bundes mittels Beiträgen brachte neue Verzerrungen mit sich. Bauern in Berggebieten, denen der Weg zur Milchannahmestelle zu weit war, verwendeten die von ihren Kühen produzierte Milch zur Mast der eigenen Kälber oder derjenigen der Talbauern. Gegenüber den milchverkaufenden Bauern, die von den Preisstützungen profitierten, waren die Landwirte in Berggebieten strukturell benachteiligt. Diesem Umstand wurde dadurch begegnet, dass den Bergbauern 1959 erstmals Beiträge aus Bundesgeldern zugesprochen wurden: die ersten Direktzahlungen.

Ausdifferenzierung von Direktzahlungen und diesbezüglichen Daten

Als Träger der Agrarpolitik befindet der Bund über Höhe und Modalitäten der Direktzahlungen (landwirtschaftliche Beiträge). Die Zuteilung der Gelder erfolgt in enger Zusammenarbeit mit den kantonalen Behörden und den Landwirten selber.⁷ Es wurde dazu ein Melde- und Beitragssystem geschaffen, das alle direkt betroffenen Akteure involviert. Tiere werden dabei bis heute in so genannte „Grossvieheinheiten“ umgerechnet. Eine Kuh zählt beispielsweise als

.....
4 StAGR X 3 x. – Auch die Zoneneinteilung der Landwirtschaftsflächen in Berggebieten im Zuge der Einführung der Direktzahlungen wurde im Landwirtschaftlichen Produktionskataster verzeichnet. Da mit der Zoneneinteilung eines Gebiets Beitragsansprüche verbunden waren (und sind), führte diese nicht selten zu Rechtsstreitigkeiten.

5 Der Paritätslohn war schwierig zu berechnen und politisch umstritten. Das Ziel paritätischer Löhne in Landwirtschaft und Industrie wurde den gängigsten Berechnungsmethoden zufolge nicht erreicht. 1998 wurde die Richtgrösse ersatzlos aus der Gesetzgebung gestrichen. Dennoch attestiert ihr der entsprechende Artikel im Historischen Lexikon der Schweiz, „doch viel zum sozialen Ausgleich zwischen den Wirtschaftssektoren“ beigetragen zu haben.

6 Verordnung vom 21. Dezember 1953 über wirtschaftliche Bestimmungen des Landwirtschaftsgesetzes (Allgemeine Landwirtschafts-Verordnung, ALV, SR 916.01).

7 Für eine Darstellung dieses Zusammenspiels siehe die Tagungspräsentation, Folie 4.

1 Grossvieheinheit, ein Pferd als 0.7.⁸ Zu Beginn gestalteten sich Melde- und Beitragsfluss wie folgt:

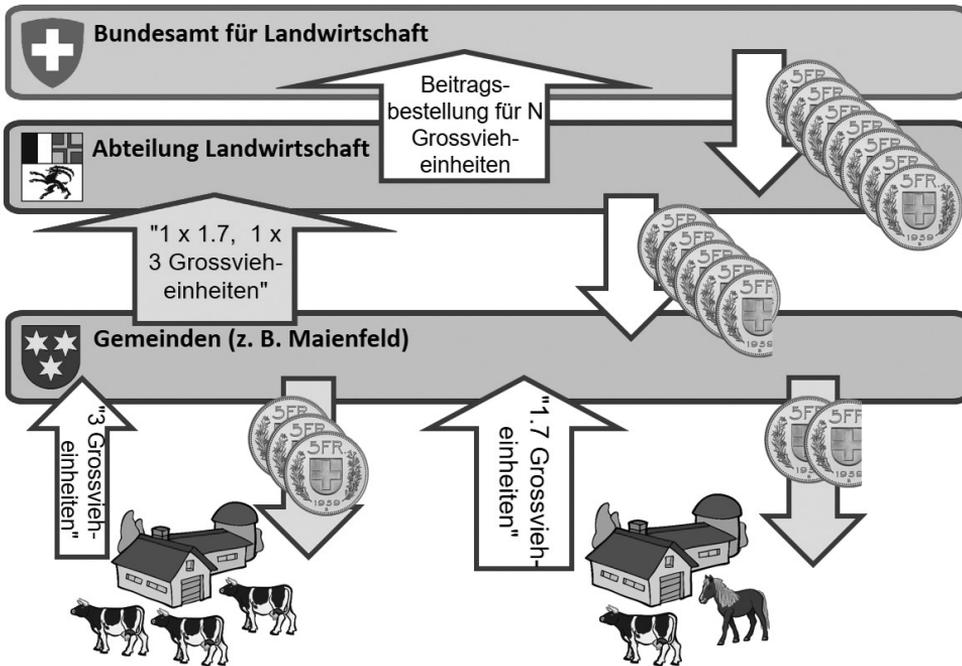


Abb. 1: Eigene Darstellung

Die Erhebung der beitragsberechtigten Tiere war ursprünglich Sache der Gemeinden. Diese beauftragten bestimmte Personen, die benötigten Zahlen bei den einzelnen Landwirten zu erheben. Nachdem die Gemeinden die Zahlen aller Betriebe gesammelt hatten, meldeten sie diese dem kantonalen Amt für Landwirtschaft, heute Amt für Landwirtschaft und Geoinformation (ALG), welches die daraus resultierenden Beiträge berechnete und beim Bund bestellte. Nach einer Kontrolle aufgrund der zur Verfügung stehenden Daten zahlte der Bund den Kantonen die Beiträge aus, worauf diese an die Gemeinden und schliesslich an die einzelnen Landwirtschaftsbetriebe verteilt wurden. Die Direktzahlungen der Bündner Landwirtschaftsbetriebe sind im Staatsarchiv Graubünden durch eine fast lückenlose Datenreihe dokumentiert.⁹ Es liegen insbesondere die Angaben gemäss den in der Grafik grau markierten Pfeilen vor. Pro Gemeinde und Jahr gibt es ein Dokument, welches die Grossvieheinheiten nach Betrieb sowie die beantragten Beiträge auflistet:

8 Die aktuellen Faktoren zur Umrechnung des Tierbestands in Grossvieheinheiten sind im Anhang zur Verordnung über landwirtschaftliche Begriffe und die Anerkennung von Betriebsformen (Landwirtschaftliche Begriffsverordnung, LBV, SR 910.91) vom 7. Dezember 1998 zu finden.

9 StAGR C2. Die Daten sind ab 1960 vorhanden, wobei für das Jahr 2000 eine Lücke vorliegt. Letztere entstand, da die entsprechende Datenbank zum Zeitpunkt der Übernahme nicht mehr mit zumutbarem Aufwand abgreifbar war.

Kostenbeiträge an Berggebiete			149	Auszahlungsliste Nr. : 1		Distinta di pagamento No : 1		B
Contributi alle spese nelle zone di montagna					Jahr : 1962			
Gemeinde : M a i e n f e l d			Kanton : Graubünden					
Comune :			Cantone :					
Name Cognome	Vorname Nome	Ortsbezeichnung (Postadresse) Denominazione locale (indirizzo postale)	Reg. Nr. der Betriebs- karte No. Reg. della scheda az.	Anzahl Grossvieheinheiten Numero di unità best. gr.		Kostenbeiträge Contributi alle spese		
				Zone 1 Zona 1	Zone 2 Zona 2	Zone 1 Zona 1	Zone 2 Zona 2	
			1	5		200.-		
			2	5		200.-		
			3	5		200.-		
			4	5		200.-		
T o t a l					20,0	800.-		

Abb. 2: Beispiel aus dem Jahr 1962. StAGR C2.3

Aus diesem frühen Formular wird bereits eine gewisse Ausdifferenzierung ersichtlich: das Berggebiet wurde in zwei Zonen, unterschieden durch die Steilheit des Geländes, aufgeteilt. Die Landwirte gaben die Grossvieheinheit pro bewirtschafteter Zone an. Für Tiere in der steileren und somit schwerer zu bewirtschaftenden Zone 2 erhielten sie höhere Beiträge als für solche in der flacheren Zone 1. In Maienfeld gab es 1960 rund 120 Viehbesitzer, wobei aber nur vier Betriebe im Berggebiet lagen und somit Direktzahlungen beantragen konnten.¹⁰

Zu einem weiteren Ausbau der Direktzahlungen kam es in den 1980er-Jahren durch den Umstand, dass sich die Bauern in Tal- und Hügelregionen von der – auf die Berggebiete zugeschnittenen – Direktzahlungspolitik benachteiligt fühlten. So wurden 1981 zusätzlich so genannte „Flächenbeiträge“ für Ackerbaubetriebe eingeführt, die schwer zugängliche Flächen bewirtschafteten.

Im Gleichschritt mit dem zunehmenden Ausbau und der wachsenden Ausdifferenzierung der landwirtschaftlichen Beiträge wurden auch die von den Behörden gepflegten Datenreihen immer umfangreicher und detaillierter. Ab Ende der 1980er-/Anfang der 1990er-Jahre wurde in der Agrarpolitik ein Paradigmenwechsel eingeleitet, indem Natur- und Landschaftsschutz einen höheren Stellenwert erhielten. Die Existenzsicherung der Bauernfamilien an sich wurde hingegen nicht mehr als Ziel gesehen. 1996 wurde der neue Art. 104 der Bundesverfassung vom Stimmvolk angenommen. Darin werden so genannte „gemeinwirtschaftliche Leistungen“ definiert, die die Landwirtschaft zu erbringen hat:

10 StAGR X 6 s 5.

„Der Bund sorgt dafür, dass die Landwirtschaft durch eine nachhaltige und auf den Markt ausgerichtete Produktion einen wesentlichen Beitrag leistet zur:

- a. sicheren Versorgung der Bevölkerung;
- b. Erhaltung der natürlichen Lebensgrundlagen und zur Pflege der Kulturlandschaft;
- c. dezentralen Besiedlung des Landes.“¹¹

Die Direktzahlungen etablierten sich als wichtigstes Instrument der Landwirtschaftsförderung. Art. 104 bestimmt weiter, dass der Bund „das bäuerliche Einkommen durch Direktzahlungen zur Erzielung eines angemessenen Entgelts für die erbrachten Leistungen [ergänzt], unter der Voraussetzung eines ökologischen Leistungsnachweises.“

Diese Neuerungen schlugen sich deutlich in der Direktzahlungspraxis nieder, wie das Beispiel eines Datenblatts pro 1995 zeigt:

LANDWIRTSCHAFTSAMT GRAUBÜNDEN 7000 CHUR		KOSTENBEITRAG / GEMEINDETOTAL ABRECHNUNG: 01 / 1995							SETTE 41 LISTE LMAP24-6 DATUM 11.7.1995					
ZO	BEITRAGSBERECHTIGTE	GVE					BEITRAG ANZ.	ABZUG ANZ.	ABZUG ANZ.					
	KUEHE RINDVIEH SCHWEINE	PFERDE ZIEGEN SCHAFEN	TOTAL	BRUTTO BETR.	RAUFUTTER BETR.	EINK/VERM. BETR.	BEITRAG							
GEMEINDE: 3953 MAIENFELD														
TO	261.25	17.59	0.00	25.90	2.54	95.07	37.94	33,306	32	1,877	1	0	0	31,429
BUNDESBEITRAG													8	31,429

Abb. 3: StAGR C2.36

Für ihre Grossvieheinheiten konnten die Betriebe weiterhin zonenabhängige (dafür steht das „ZO“ oben links) Direktzahlungen beantragen. Jedoch wurden Abzüge gemacht, wenn die Tiere zu wenig oder gar kein Raufutter¹² erhielten. Bei dem Datenblatt handelt es sich im Übrigen um ein klassisches Dokument aus der „EDV-Zeit“: das kantonale Landwirtschaftsamt verwaltete die Direktzahlungsdaten in einer Datenbank; für den Austausch mit dem Bund und zum Zweck der Archivierung wurden jedoch Ausdrucke erstellt.

Auch am nächsten Beispiel werden sowohl hinsichtlich Direktzahlungspraxis als auch Datenhaltung und -bearbeitung Veränderungen sichtbar. Es handelt sich um einen Auszug aus der Datenbank „Apoma“ der Abteilung Landwirtschaft des kantonalen Amtes für Landwirtschaft und Geoinformation (ALG) für das Jahr 1999. Der Datenaustausch funktionierte dabei mittels CSV-Tabellen, die auch in dieser Form dem Staatsarchiv abgeliefert wurden.

11 Bundesverfassung der Schweizerischen Eidgenossenschaft (SR 101) vom 18. April 1999.

12 Die Fütterung mit Raufutter gilt als besonders artgerecht und ökologisch, da Raufutter – im Vergleich zu importierten Sojaprodukten usw. – aus heimischen erneuerbaren Ressourcen wie Gras hergestellt wird.

Gmd Nr.	Name	Ort	BIO	Öln	Label	BTS	RAUS	401	402	403	404
			1	ON	N	J		0	0	0	0
			1	ON	N	J		0	0	0	0
			0	ON	N	N		0	0	0	0
			1	ON	J	J		0	0	0	0
			1	ON	N	J		0	0	0	0
			1	ON	N	J		0	0	0	0
			0	ON	N	N		0	0	0	0
			1	ON	J	J		210	0	0	0
			1	ON	N	J		160	133	0	0
			0	ON	N	N		0	0	0	0

Abb. 4: StAGR C12.1

Der verstärkte Fokus auf den Natur- und Landschaftsschutz spiegelt sich darin klar wider: Der Betrieb auf der achten Zeile im obigen Beispiel beantragte z. B. Direktzahlungen für 210 Aren „Extensive Wiese mit seltenen Blumen“ (Code 401). Die Ackerflächen werden in Aren oder Hektaren gemessen, die angebauten Feldfrüchte mit einem Code erfasst.¹³ Der Ausbau der Direktzahlungen hatte zur Folge, dass 1999 in Maienfeld 50 von insgesamt rund 60¹⁴ Landwirtschaftsbetrieben Direktzahlungen beanspruchten.¹⁵

Die Einführung der Flächenbeiträge ab 1981 und deren Erweiterung in den 1990er-Jahren hatte auch einen Ausbau der Strukturen für die Datenerhebung zur Folge. Sie präsentieren sich aktuell wie folgt:

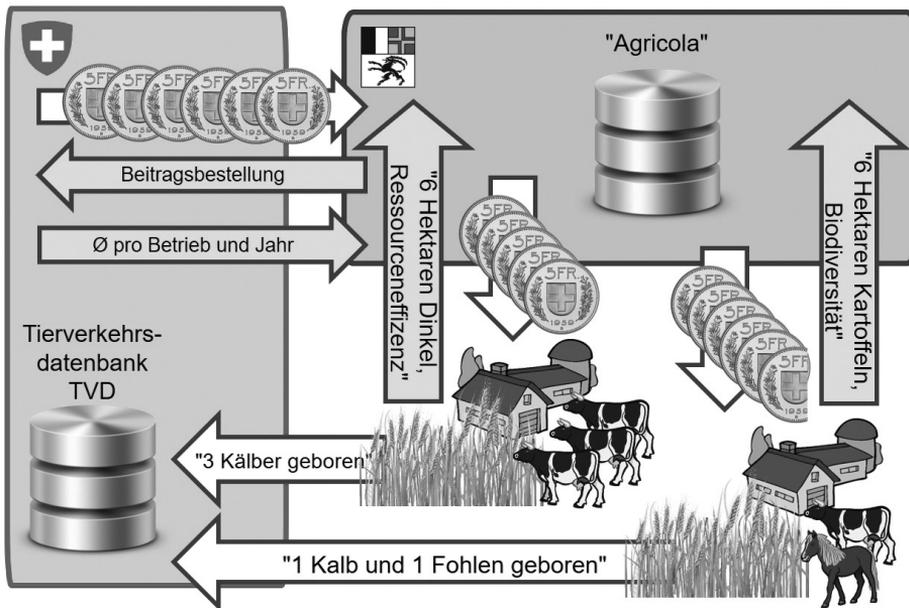


Abb. 5: Eigene Darstellung

13 1 Are (a) = 100 Quadratmeter, 1 Hektare (ha) = 10'000 Quadratmeter (bzw. 100 a).

14 StAGR C II 12 a 2, Graubünden in Zahlen 1999.

15 Vgl. Abb. 3: 1995 beantragten 32 Landwirtschaftsbetriebe Beiträge.

Tierzahlen werden nicht mehr von Gemeindebeauftragten pro Gemeinde erhoben und gemeldet. Vielmehr werden die Geburten und Todesfälle bzw. die Schlachtung von Tieren von den Landwirten selber in der Tierverkehrsdatenbank (TVD) des Bundes erfasst.¹⁶ Eine Applikation des Bundes errechnet jährlich den durchschnittlichen Tierbestand der Betriebe übers Jahr. Diese Daten werden den kantonalen Landwirtschaftsbehörden weitergeleitet, wo sie in der Datenbank „Agricola“ gespeichert werden.¹⁷ Die Erhebung der Flächendaten erfolgt ebenfalls durch Selbstdeklaration der Betriebe über ein kantonales Agricola-Interface. Dabei müssen die Landwirte auch angeben, an welchen Beitragsprogrammen sie mitmachen wollen (im Beispiel Beiträge für Ressourceneffizienz beziehungsweise Biodiversität).

Aus Flächen- und Tierdaten errechnet das ALG jährlich die Beiträge, die den einzelnen Betrieben zustehen und bestellt die Gelder beim Bundesamt für Landwirtschaft (BLW). Nach einer Gegenprüfung anhand der gelieferten Daten überweist das BLW dem ALG die bestellten Gelder, worauf sie an die einzelnen Betriebe ausbezahlt werden. Gegenwärtig beziehen allein die Bündner Landwirtschaftsbetriebe jährlich rund 225 Millionen Franken Direktzahlungen.¹⁸

Die Agricola-Datenbank beinhaltet folglich alle für das Funktionieren des Direktzahlungssystems notwendigen Daten sowie viele weitere landwirtschaftliche Strukturdaten. Die kantonalen Behörden liefern diese Daten jährlich in vereinheitlichter Form dem BLW, wobei die Datenhoheit beim Kanton bleibt. Seit 2004 übernimmt das Staatsarchiv Graubünden diese Daten und setzt damit die Datenreihe fort, die die Beitragszahlungen an Bündner Bauernbetriebe seit 1960 dokumentiert.

Ein Agricola-Export gemäss den Standards des BLW (in dieser Form werden die Daten auch vom Staatsarchiv Graubünden übernommen) teilt sich in vier XML-Dateien:

- Ein Register der Betriebe und Landwirte
- Eine Datei „EcoEtho“ mit den Anmeldedaten für die Beitragsprogramme
- Eine Datei „Structure“ mit aggregierten Daten zur Berechnung der Direktzahlungen
- Eine Datei „Payment“ mit Angaben zu den effektiv ausbezahlten Direktzahlungen für Tiere und Flächen sowie allfälligen Korrekturen

Ein Beispiel aus der „Payment“-Datei von 2015:¹⁹

```
<bdQualityPerCulture> Biodiversitäts-Qualitätsbeitrag pro Kulturcode
  <C_Kultur>611</C_Kultur> Kulturcode (611 = extensiv bewirtschaftete
  Wiesen)
  <F_Menge>14400</F_Menge> Flächen Menge (14'400 Quadratmeter)
  <B_Total>945.00</B_Total> Berechneter Qualitätsbeitrag
</bdQualityPerCulture>
```

16 Auch andere für Nutztiere zuständige Instanzen wie Tierärzte und Schlachthäuser sind zur Pflege der TVD verpflichtet.

17 Agricola wird von den meisten Ostschweizer Kantonen sowie von Zürich und dem Aargau benützt. Vier andere Applikationen werden in den übrigen Kantonen verwendet.

18 Der Gesamtbetrag der Direktzahlungen belief sich 2018 auf 2.8 Milliarden Franken, <https://www.agrarbericht.ch/de/politik/direktzahlungen/finanzielle-mittel-fuer-direktzahlungen> [Abruf 24.4.2019].

19 StAGR 2016/035.

Dokumentation der landwirtschaftlichen Daten

Form und Inhalt der landwirtschaftlichen Daten, die zwischen Bundes- und Kantonsbehörden ausgetauscht werden, gibt – wie oben erwähnt – das BLW vor. Dieses gibt zu den Daten jedes Jahr – nebst weiteren Dokumentationsunterlagen – aktualisierte so genannte „Merkmalskataloge“ heraus, die unter anderem die zu liefernden XML-Dateien technisch beschreiben und die aktuellen Kulturcodes enthalten. Z. B. wurde 2018 der Anbau von Senf beitragsberechtigt, was die Einführung des Kulturcodes „573 Senf“ mit sich brachte.

Damit die landwirtschaftlichen Daten auch nach der Übernahme ins Archiv verständlich bleiben, ist es notwendig, die Merkmalskataloge ebenfalls zu übernehmen (als PDF). Zur Erhöhung des Evidenzwerts werden zudem die Formulare für die Datenerhebung mit übernommen. Die meisten Landwirte benutzen mittlerweile statt den Formularen das Agricola-Interface; die Formulare enthalten aber dieselben Felder und sind somit eine einfache Form der dauerhaften Aufbewahrung der entsprechenden Informationen. Zur Erleichterung der künftigen (automatisierten) Benutzung werden nebst den vier übernommenen XML-Dateien der landwirtschaftlichen Daten auch die zugehörigen XML-Schemata dauerhaft aufbewahrt. Dies wäre an sich zum Verständnis der XML-Dateien nicht zwingend notwendig, da alle benötigten Informationen aus den Merkmalskatalogen hervorgehen.

Die vom BLW herausgegebene Dokumentation²⁰ enthält insgesamt die nachfolgend beschriebenen Unterlagen. Das Beispiel zeigt die Dokumentation des Beitragsjahrs 2018.

Steckbrief Landwirtschaftliche Struktur-erhebung	Der Steckbrief beschreibt in knapper Form die Landwirtschaftliche Struktur-erhebung, ihre Methodik und gesetzlichen Grundlagen.
Merkmalskatalog Anmeldungen (EcoE-tho) 2018	Dieser Merkmalskatalog beschreibt die Übermittlung der Anmeldungsdaten für die Direktzahlungsarten, den Ökologischen Leistungsnachweis (ÖLN) und die ÖLN-Kontrollstelle (Art. 97 DZV) durch die Kantone an das BLW im XML-Format. ²⁶ Er beschreibt die fachliche Bedeutung der XML-Elemente und kommt für die Datenerhebung im Herbst 2017 (Beitragsjahr 2018) zum Tragen.
Merkmalskatalog Payment 2018	Dieser Merkmalskatalog beschreibt die Form, in welcher die Beitragsdaten von den Kantonen ans BLW (Agrarpolitisches Informationssystem (AGIS)) und vom Beitragsberechnungsservice (BBS) des BLW an die Kantone geliefert werden.

20 Der Steckbrief bildet eine Ausnahme: Dieser wird nicht vom BLW, sondern vom Bundesamt für Statistik (BFS) herausgegeben.

21 Verordnung über die Direktzahlungen an die Landwirtschaft (Direktzahlungsverordnung, DZV, SR 910.13) vom 23. Oktober 2013.

	Die Hierarchie der XML-Elemente (die Baumstruktur der XML Datei) sowie der Inhalt und die Eigenschaften jedes Elementes werden definiert
Merkmalkatalog Strukturdaten 2018	Dieser Merkmalkatalog enthält Richtlinien für die Übertragung von landwirtschaftlichen Strukturdaten von den Kantonen ans BLW im XML-Format. Er legt den Umfang und das Format der zu liefernden Daten fest. Dazu werden die Hierarchie der XML-Elemente (die Baumstruktur der XML-Datei) sowie der Inhalt und die Eigenschaften der Elemente festgelegt.
Merkmalkatalog Registerdaten 2018	Dieser Merkmalkatalog beschreibt, wie die Daten zu Betriebsformen und Personen (Registerdaten) von den Kantonen zu erfassen und ans BLW zu übermitteln sind. Die Verwaltungsregeln für die Registerdaten (Teil I) bieten einen Überblick über die Richtlinien zum Inhalt und zur Verwaltung der kantonalen Register mit landwirtschaftlichen Betriebsformen und Personen (Bewirtschaftern). Die Merkmalsbeschreibung (Teil II) legt den Umfang und das Format der zu liefernden Daten fest. Dazu werden in diesem Teil die Elemente der zugrunde liegenden XML-Schemata fachlich konkretisiert.
Formular Allgemeine Angaben 2018	Das Formular dient im Rahmen der landwirtschaftlichen Strukturhebung der Erfassung von allgemeinen Angaben durch Selbstdeklaration der Betriebe.
Formular Flächenerhebung 2018	Das Formular dient im Rahmen der landwirtschaftlichen Strukturhebung der Erfassung von Flächendaten (Grösse, Nutzung, Hanglage, Biodiversitätsbeiträge usw.) durch Selbstdeklaration der Betriebe.
Formular Tiererhebung 2018	Das Formular dient im Rahmen der landwirtschaftlichen Strukturhebung der Erfassung von Tierbeständen durch Selbstdeklaration der Betriebe.
Formular Bienenstände 2018	Das Formular dient im Rahmen der landwirtschaftlichen Strukturhebung der Erfassung von Bienenständen durch Selbstdeklaration der Imker.
Formular Aquakulturbetriebe 2018	Das Formular dient im Rahmen der landwirtschaftlichen Strukturhebung der Erfassung von Arten, Art der Haltung und Produktionsform durch Selbstdeklaration der Aquakulturbetriebe.
EcoEtho 2018	XML-Schema (XSD) für die EcoEtho-Datei 2018.
Register 2018	XML-Schema (XSD) für die Register-Datei 2018.
Structure 2018	XML-Schema (XSD) für die Structure-Datei 2018.
Payment 2018	XML-Schema (XSD) für die Payment-Datei 2018.

Ausblick

Alle vier Jahre legt das Bundesparlament die agrarpolitische Stossrichtung neu fest. Mittels Referendum und Initiative kann auch das Stimmvolk darauf Einfluss nehmen. Aktuell wird nach den – in zahlreichen Erlassen festgeschriebenen – Bestimmungen der „Agrarpolitik 18-21“ gearbeitet. Die „Agrarpolitik 22+“ ist in Vorbereitung; die Vernehmlassung ist abgeschlossen und wird derzeit von den Bundesbehörden ausgewertet.²² Wie oben erläutert, bedingte die konkrete Ausgestaltung der Agrarpolitik in der Vergangenheit stets auch Anpassungen im Bereich der zu erhebenden bzw. anfallenden Daten. Diese Anpassungen können inhaltlicher (beispielsweise infolge Schaffung neuer Beitragsprogramme) oder technischer (etwa durch Weiterentwicklung der Datenaustauschinfrastruktur) Natur sein. Das Staatsarchiv Graubünden beobachtet deshalb die politischen Entwicklungen auch auf Bundesebene und pflegt auf kantonaler Ebene insbesondere den Kontakt mit dem ALG. So wird sichergestellt, dass die archivierten Datenreihen nicht abreißen, sondern im Gegenteil auch die Veränderungen in der Agrarpolitik und deren Umsetzung im Kanton Graubünden weiterhin mit dokumentiert werden.

Wie oben umrissen, erfährt die Art und Weise des Austauschs landwirtschaftlicher Daten auch während laufenden agrarpolitischen Perioden kleinere Änderungen. Solche gilt es bei der Archivierung nicht nur zu berücksichtigen, sie stellen bisweilen auch eine Chance dar, den gesetzlichen Auftrag des Staatsarchivs noch besser wahrzunehmen.²³ Ein Beispiel dafür ist der Umstand, dass das BLW von den kantonalen Behörden ab dem Beitragsjahr 2019 zusätzlich zu den landwirtschaftlichen Daten im XML-Format auch Geodaten im INTERLIS-Format verlangt. Diese landwirtschaftlichen GIS-Daten stellen die meisten der auch in den XML-Daten enthaltenen Informationen geographisch dar.²⁴ Durch die Übernahme der Geodaten ins Archiv wird die stark räumlich ausgerichtete Arbeitsweise des ALG dokumentiert und der Forschung steht eine attraktive und vielseitig nutzbare zusätzliche Darstellungsform von Informationen aus dem Bereich Landwirtschaft zur Verfügung.

Das beim Austausch von Geobasisdaten zwischen Bundes- und Kantonsbehörden verwendete Format „INTERLIS“ wurde bereits in einem Projekt der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) als archivtauglich bewiesen.²⁵ Abschliessend folgt zur Veranschauli-

.....
22 Für weitere Informationen zur aktuellen Agrarpolitik siehe <https://www.blw.admin.ch/blw/de/home/politik/agrarpolitik.html> [Abruf 24.4.2019].

23 Vgl. Gesetz über die Aktenführung und Archivierung (GAA, BR 490.000) vom 28. August 2015. Im vorliegenden Zusammenhang besonders zu erwähnen: Art. 2, Abs. 1, Lit. b) „Nachvollziehbarkeit und Dokumentierung des Handelns der Behörden“ und e) „Unterstützung der Forschung“.

24 „GIS“ = „Geographisches Informationssystem“. Die GIS-Daten zum Direktzahlungsvollzug in Graubünden sind hier zu finden: <https://geo.gr.ch/gis-tools/gdds/inventar/detailinventar.php> [Abruf 24.4.2019]. – Ein wesentlicher Unterschied gegenüber den XML-Daten ist, dass die Frankenbeträge der Direktzahlungen in den GIS-Daten nicht enthalten sind.

25 Projekt 14-016 AV. Der Schlussbericht ist hier zu finden: https://kost-ceco.ch/cms/14-016_de.html [Abruf 24.4.2019]. – Zur Abgrenzung Geodaten-Geobasisdaten siehe <https://www.geo.admin.ch/de/geoinformation-schweiz/geobasisdaten/faq/geobasisdaten-des-Bundesrechts.html> [24. April 2019].

chung ein Beispiel aus einem zu Testzwecken vom ALG übernommenen Datensatz der Gemeinde Maienfeld.



Abb. 6: Luftbildfotografie

Vor dem Hintergrund einer Luftbildfotografie werden hier die Perimeter der landwirtschaftlichen Nutzfläche (graue Linien) sowie Biodiversitäts-Förderflächen (dunkel- und hellgrau gepunktete Bereiche) angezeigt. Das Datenobjekt für die auf Grundstück Nr. 676 zur Anwendung gebrachte Direktzahlungsart wird im entsprechenden INTERLIS-File folgendermassen beschrieben: „OBJE 3186 ANUXXXXXXX 3953/_X/XXX 676 3953 717 0 113.00 @“. ²⁶ Dabei bedeuten:

3186	die vom ALG vergebene Objektnummer
676	die Grundstücksnummer (676)
3953	die BFS-Gemeindenummer (3953 = Maienfeld)
717	der Kulturcode (717 = Rebflächen mit natürlicher Artenvielfalt)
113.00	die Fläche in Aren (113 Aren = 11'300 Quadratmeter)

Auf Parzelle Nummer 676 werden somit Beiträge für 113 Aren Rebfläche mit natürlicher Artenvielfalt beantragt. In der vorliegenden GIS-Darstellung er-

.....
²⁶ Der (leicht gekürzte) Auszug aus dem INTERLIS-File wurde durch die Autoren mit „XXX“ anonymisiert.

scheint die entsprechende Fläche sowohl gelb als auch blau gepunktet (es liessen sich auch beliebige andere Darstellungsformen wählen). Rebflächen mit natürlicher Artenvielfalt gelten als Biodiversitäts-Förderflächen der Qualitätsstufe II (hellgraue Punkte) und erhalten Vernetzungsbeiträge (dukelgraue Punkte). Zum Vergleich: auf Parzelle Nr. 684 befindet sich eine extensiv bewirtschaftete Wiese der Qualitätsstufe I (in dieser Darstellung keine besondere Kennzeichnung). Für extensiv bewirtschaftete Wiesen können Vernetzungsbeiträge beantragt werden (dunkelgraue Punkte).

Die Grundstücksnummer wird von der Amtlichen Vermessung vergeben; die betreffenden Eigentumsverhältnisse werden von den Grundbuchämtern beurkundet und registriert. Die Datenhoheit der als Hintergrund verwendeten Luftbildfotografie liegt beim Bundesamt für Landestopographie swisstopo. Die für die Arbeit der Behörden (wie auch von Privaten) bereits stark genutzte Verknüpfung von Geodaten aus verschiedenen Quellen soll auch im Bereich der Archivierung vermehrt fruchtbar gemacht werden – in Sachen Übernahme von landwirtschaftlichen Daten aus Fachapplikationen gibt es also weiterhin viel Spannendes zu tun.

Summary

Since the creation of the modern Swiss Confederation in 1848, to fulfil their increasing amount of tasks, the (federal) agencies have been collecting data that became ever more detailed. This also applies for the agricultural sector, which over time has been confronted with manifold challenges from securing the country's food supply during times of crisis to increasing international competition and, since the late 1980s, calls for a more ecological and sustainable production. In 1959/60, the pegging of milk prices as a means of agricultural policy has led to the introduction of so-called „Direktzahlungen“ (literally „direct payments“) for cattle-breeders in mountainous regions.²⁷ This direct payments have since been extended to crop farming in the lowlands. The State's Archives of Graubünden hold a nearly continuous series of survey and financial data documenting in detail the direct payments made in the canton. While the series starts as a collection of printed forms, the current data format used for information exchange between government agencies and for archiving is XML. A main opportunity and challenge for the State's Archives is to keep up with the changing political and technical framework of agricultural administration. One next step in this direction will be the accession of agricultural geodata.

.....
 27 In mountainous regions, the distance from the farm to the next dairy cooperative is often too far for economically rewarding transportation of milk. Instead, the farmers in this areas use the milk of their cows for cattle breeding. In comparison with dairy farmers, they were disadvantaged by the introduction of price subsidies for milk.

DCAT-AP.DE. DER METADATENSTANDARD ZUM AUSTAUSCH OFFENER VERWALTUNGSDATEN

Paul Flamme

Bund, Länder und Kommunen verfügen über eine enorme Anzahl von Datensammlungen, die als öffentliche Güter frei zugänglich zu machen sind. *DCAT-AP.de* ist der deutsche Metadatenstandard für offene Verwaltungsdaten (Open Government Data). Er wurde im Juni 2018 vom IT-Planungsrat als formaler Austauschstandard für offene allgemeine Verwaltungsdaten festgelegt.¹

Offene Verwaltungsdaten, Open Data, Big Data und Künstliche Intelligenz sind der Stoff aus dem in Zukunft zunehmend gesellschaftliche Zusammenhänge sichtbar gemacht werden. In 2013 verabschiedeten die G8-Regierungschefs die *Open Data Charta*. 70 Staaten sind der inzwischen der Verpflichtung beigetreten, um durch eine breite Öffnung die Daten der Verwaltung innovativ weiterverwenden zu können.² Die Vorlage eines *Nationalen Aktionsplanes der Bundesregierung zur Umsetzung der Open-Data-Charta der G8*³ im Jahre 2014 ist Teil dieser Verpflichtung. Damit beginnt in Deutschland – quantitativ, aber auch qualitativ – eine deutliche Dynamisierung der Veröffentlichung offener Verwaltungsdaten. In Umsetzung der Verabredungen der G8-Staaten kündigte die Bundesregierung die Einrichtung eines zentralen OpenData-Portals für Bund, Länder und Kommunen sowie die Etablierung eines auf Interoperabilität basierenden Metadatenstandards für offene Daten an.⁴

Die Archivarinnen und Archivare sollten diese Entwicklung offensiv begleiten, verstehen sie sich doch schon jetzt auch als Informationsdienstleister und sehen in der Förderung der Transparenz staatlichen Handelns eine ihrer Aufgaben.⁵ Zuverlässige und relevante Informationen zu extrahieren, bereitzustellen

-
- 1 Vgl. Unterlagen und Entscheidungen Standardisierungsbedarf „Metadatenstruktur für offene Verwaltungsdaten“ der 26. Sitzung des IT-Planungsrates, URL: https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2018/Sitzung_26.html?pos=9 [Abruf 15. 5. 2019].
 - 2 Vgl. <https://opendatacharter.net/> [Abruf 15. 5. 2019].
 - 3 Vgl. https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2014/aktionsplan-open-data.pdf?__blob=publicationFile&v=1 [Abruf 15. 5. 2019].
 - 4 Wie Anm. 1, S. 7.
 - 5 Zum Spannungsverhältnis Informationsfreiheitrecht - Archivrecht vgl. Lehnstaedt, Stephen/Stemmer, Bastian: Über die Einsicht in staatliche Dokumente vor deren Archivierung. In: *Archivar* 66 (2013), S. 46–48; Wiech, Martina: Informationsfreiheit. Eine Erwiderung aus archivischer Sicht zum Beitrag von Stephan Lehnstaedt und Bastian Stemmer. In: *Archivar*

und dauerhaft zu erhalten wird mit zunehmender Öffnung noch arbeitender Registraturen kein Monopol der Archive bleiben.⁶

Kenntnisse über Funktionsweise und Ausgestaltung von Metadaten gehören zur Grundausstattung der in Archiven Tätigen. Für die Übernahme elektronischer Unterlagen in ein digitales Archiv ist die Nutzung und Parametrierung einschlägiger Metadatenstandards wie Premis, METS und Dublin Core unverzichtbar. Da jede Übernahme in ein elektronisches Archiv ihre Besonderheiten hat und jeder Datentyp neue erhebliche Aufwände verursacht, sollten neue Standards für die Beschreibung und den Austausch von Verwaltungsdaten daraufhin geprüft werden, ob sie sich für die Übernahme in ein elektronisches Archiv eignen.

DCAT => DCAT-AP => DCAT-AP.de

„Offene Daten sind Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können – maximal eingeschränkt durch Pflichten zur Quellennennung und Urheberrechte.“⁷ So lautet die Definition des Bundesministeriums für Verkehr und digitale Infrastruktur in Anlehnung an die *Open Knowledge Foundation*. Mit Begrenzung auf die öffentliche Verwaltung spricht der IT-Planungsrat von „Datenbestände des öffentlichen Sektors, die von Staat und Verwaltung im Interesse der Allgemeinheit ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden.“⁸

Seit der Selbstverpflichtung der G8-Staaten steht die Etablierung einer einheitlichen „Metadatenstruktur für offene Verwaltungsdaten“ auf der Standardisierungsagenda des IT-Planungsrates. Die Fachgruppe GovData, das ist ein Bund- / Ländergremium, das das gleichnamige Portal für die offenen Daten von Bund, Ländern und Kommunen betreibt, hat im November 2016 beschlossen, dass als deutscher Metadatenstandard für offene Verwaltungsdaten eine DCAT-AP-konforme Ableitung zu erstellt sei.

Mit DCAT (*Data Catalog Vocabulary*) wurde ein RDF-Vokabular mit dem Ziel geschaffen, die Interoperabilität zwischen Online-Datenkatalogen im Web zu gewährleisten. Es handelt sich um eine Spezifikation, welche unter Verantwor-

.....
66 (2013), S. 49–50; Kuppe, Stefan/Schäfer, Udo: Das Transparenzportal Hamburg. Open Government Data als Angebot auch an die historische Forschung. In: Recherche und Weiterverarbeitung. Beiträge einer Sektion auf dem 51. Deutschen Historikertag 2016 in Hamburg, Stuttgart 2017, S. 52-62; Flamme, Paul: Eine neue Aufgabe für ein staatliches Archiv: das Transparenzportal Hamburg. In: VdA (Hrsg.), Transformation ins Digitale. 85. Deutscher Archivtag in Karlsruhe. Tagungsdokumentation, Fulda 2017, S. 29 – 42.

6 Trotz der damit verbundenen Synergien durch Kompetenzbündelungen und im technischen Bereich, ist wohl nicht davon auszugehen, dass der Betrieb der OpenData-Portale bzw. Transparenzportale flächendeckend bei den staatlichen Archiven angesiedelt wird. Zu Transparenzportal und Archiv siehe Flamme, Transparenzportal, wie Anm. 5, S. 30 ff.

7 Bundesministerium für Verkehr und digitale Infrastruktur (Hrsg.), Positionspapier Open Data des BMVI, Version 1.0, Berlin Mai 2017, S. 5.

8 Bericht zum Standardisierungsbedarf des IT-PLR „Metadatenstruktur für offene Verwaltungsdaten“, 05.03.2018, S. 20, URL: https://www.it-planungsrat.de/SharedDocs/Sitzungen/DE/2018/Sitzung_26.html?pos=9 [Abruf 15. 5. 2019].

tung der Arbeitsgruppe *Government Linked Data Working Group* beim *World Wide Web Consortium (W3C)* entwickelt und am 16. Januar 2014 vom W3C veröffentlicht wurde.⁹ DCAT setzt auf dem etablierten und bewährten Metadaten-Set *Dublin Core* auf, welches bereits im Jahr 2009 als *ISO-Standard 15836* veröffentlicht wurde.

DCAT-AP ist ein dazugehöriges europäisches Application Profile (Anwendungsprofil). Es folgt den DCAT-Empfehlungen für den Austausch von Datensatzbeschreibungen zwischen Datenportalen in Europa. Sein Schwerpunkt liegt auf einem gemeinsamen Vokabular speziell zur Beschreibung der in europäischen Datenportalen gespeicherten Datensätze des öffentlichen Sektors. Anders als beispielsweise das Containerformat METS dient DCAT.AP.de nicht dem Transport von Daten nennt aber zwingend URLs, unter der die beschriebenen Ressourcen frei zugänglich zu finden sein müssen.

DCAT-AP entstand in Zusammenarbeit mit dem *Publications Office der (of the) EU* und wird als Aktivität des *ISA² Programmes (Programm für die Interoperabilität digitaler Dienste)* des *Rats der Europäischen Union* weiter gepflegt.

Bei dem hier besprochenen Schema *DCAT-AP.de* handelt es sich dann um die deutsche Spezifikation des europäischen Profils *DCAT-AP*. Vorrangiger Zweck von *DCAT-AP.de* ist die Sicherung der Interoperabilität beim Austausch von Metadaten offener Verwaltungsdaten in Deutschland über die verschiedenen Portalebene und bei der Weiterreichung an europäische Plattformen. Die Geschäfts- und Koordinierungsstelle der Fachgruppe GovData, die Betreiberin des Datenportals für Deutschland, verantwortet die Bereitstellung und Fortschreibung des Standards *DCAT-AP.de*.¹⁰ *DCAT-AP.de* wurde von der Fachgruppe GovData gemeinsam mit dem Freistaat Sachsen und der Jinit[AG entwickelt und veröffentlicht.¹¹

Dem Grundsatz folgend, keine neuen Begriffe zu kreieren, die schon woanders definiert wurden, verwendet DCAT Klassen und Eigenschaften etablierter Vokabulare wieder. Neben den eigenen Spezifikationen DCAT (Data Catalog Vocabulary) und dcatde (German Adaption of DCAT-AP) verwendet die Ableitung universelle Standards und Spezifikationen wie:

9 Zur Entwicklung von DCAT und DCAT-AP siehe: DCAT-AP.de Spezifikation. Deutsche Adaption des „Data Catalogue Application Profile“ (DCAT-AP) für Datenportale in Europa, Version: 1.0.2, S.4.

10 DCAT-AP.de Pflegehandbuch. Pflegehandbuch für die deutsche Adaption des „Data Catalogue Application Profile“ (DCAT-AP) für Datenportale in Europa, Version 28.09.2017, S. 20ff. URL: https://www.it-planungsrat.de/SharedDocs/Downloads/DE/Entscheidungen/26_Sitzung/TOP12_Anlage1_DCAD_AP.pdf?__blob=publicationFile&v=2 [Abruf 15. 5. 2019].

11 Die technischen Ausführungen in diesem Beitrag beziehen sich neben den bereits genannten Veröffentlichungen auf die von der Fachgruppe GovData bereitgestellten einschlägigen Dokumente, insbesondere: DCAT-AP.de Spezifikation, wie Anm. 9; DCAT-AP.de Konventionenhandbuch. Technische, semantische und organisatorische Konventionen der Kooperation „GovData“, 16.03.2017; URL: <https://www.dcat-ap.de/def/> [Abruf 15. 5. 2019].

dct	Dublin Core Metadata Terms für publizierte Unterlagen
adms	Asset Description Metadata Schema / Vokabular zur Beschreibung von Interoperabilitätsbeständen
foaf	Friend of a friend / frühes RDF-Schema für Personen und Organisationen
rdfs	Ressource Description Framework Vocabulary Description Language: RDF Schema
schema	Vocabulary for structured data on the Internet
spdx	Software Package Data Exchange / Standard für Lizenzen, Copyrights für offene Software
xsd	XML Schema
vcard	Spezifikation für elektronische Visitenkarten

IT-Planungsrat beschließt Standard DCAT-AP.de

Für die Bundesrepublik Deutschland hat der IT-Planungsrat die verbindliche Nutzung von DCAT-AP.de im Juni 2018 vorgeschrieben. Beim IT-Planungsrat handelt es sich um ein Politisches Steuerungsgremium von Bund und Ländern und Kommunen in Deutschland für Informationstechnik und E-Government. Die föderale Struktur der Bundesrepublik machte es lange unmöglich, bei IT-Standards eine länderübergreifende Verbindlichkeit herzustellen. Die Einrichtung eines entsprechenden Gremiums ermöglichte in 2009 eine Grundgesetzesänderung mit der Hinzufügung des Art. 91.c. Gesetzliche Grundlage ist der „IT-Staatsvertrag“, der am 1. April 2010 in Kraft trat.¹² Dem IT-Planungsrat obliegt die Koordinierung der Zusammenarbeit von Bund und Ländern in Fragen der Informationstechnik und damit auch die Beschlussfassung über fachunabhängige und fachübergreifende IT-Interoperabilitäts- und IT-Sicherheitsstandards.¹³

Eine Verabredung zum *Standardisierungsbedarf* „Metadatenstruktur für offene Verwaltungsdaten“ traf der IT-Planungsrat am 28.06.2018. Es wurde die verbindliche Nutzung des Interoperabilitätsstandards DCAT-AP in der Ausprägung DCAT-AP.de beschlossen. Der IT-Planungsrat beauftragt die Geschäfts- und Koordinierungsstelle GovData mit dem Betrieb und der Pflege des neuen Standards.¹⁴ Zum Prozedere der Etablierung des Standards gehört auch die Festlegung: „Der Standard DCAT-AP.de ist beim Bundesarchiv, Potsdamer Stra-

12 Vgl. IT-Planungsrat. Artikel IT & Digitalpolitik; URL: <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/it-des-bundes/it-planungsrat/it-planungsrat-node.html> [Abruf 15. 5. 2019].

13 Für die Archivverbunde nicht unwichtig ist zudem die Zuständigkeit des IT-Planungsrates für die Planung und Weiterentwicklung des Verbindungsnetzes der Verwaltungen (Netze des Bundes (NdB)) für Bund, Länder und Kommunen (ehemals DOI). Beispielsweise arbeitet der Verbund Digitale Archivierung Nord (DAN) auf der Basis des NdB.

14 Vgl. Standardisierungsbedarf Metadatenstruktur, wie Anm. 1, S. 5ff.

ße 1, 56075 Koblenz, für jedermann zugänglich und archivmäßig gesichert niedergelegt. Der Standard sowie Änderungen des Standards werden vom IT-Planungsrat im Bundesanzeiger mit Herausgabedatum und Datum des Inkrafttretens verkündet.“ Die Frist zur Umsetzung dieses Beschlusses wurde für alle dem GovData-Portal zuliefernden Datenbereitsteller auf den 31. Dezember 2018 festgesetzt, für alle weiteren IT-Systeme mit dem Zeitpunkt der Änderung des jeweiligen Datenmodells des Metadatenkatalogs.

Stand Veröffentlichung offener Verwaltungsdaten mit DCAT_AP.de

Publizierte offene Verwaltungsdaten dürfen nach den einschlägigen Informationsfreiheitsgesetzen sowie dem Informationsweiterverwendungsgesetz beliebig aggregiert, gestückelt und verändert werden.¹⁵ Die Veröffentlichung der Primärdaten auf den staatlichen Informationsplattformen ist dagegen zeitlich begrenzt.¹⁶ Auf die Archive kommt bei der Anbietung dieser Unterlagen am Ende der Aufbewahrungsfrist die Aufgabe zu, die Ursprungsdaten, von denen es dann zahlreiche Ableitungen, Derivate und Ausprägungen geben kann, aufzubewahren und für notwendige Nachweise und zur Rechtssicherung bereitzuhalten. Wenn beispielsweise ein landesplanerisches Gutachten auf Geofachdaten und Baugenehmigungen fußt, die im Hamburger Transparenzportal bereitgestellt werden, bedarf es auf lange Sicht einer Sicherung der Primärdaten; sonst ist das Gutachten wertlos.

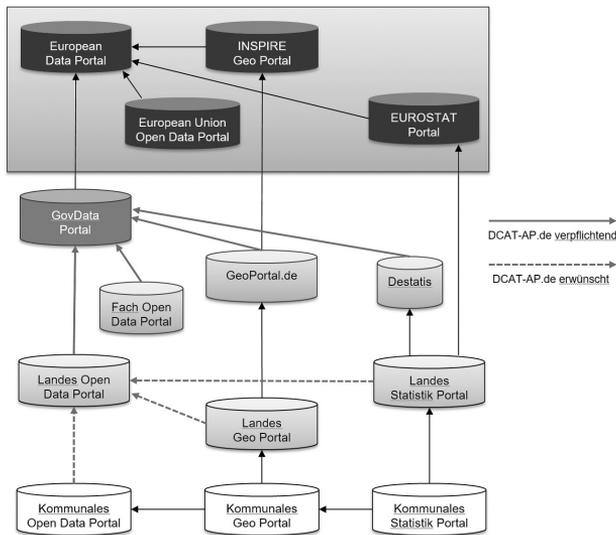


Abb. 1: Übersicht über die Nutzer von DCAT-AP.de, Quelle: Jinit[, DCAT-AP.de Pflegehandbuch¹⁷

¹⁵ Vgl. Flamme, wie Anm. 5, S. 42.

¹⁶ Nach § 10 Abs. 6 des Hamburgischen Transparenzgesetzes beispielsweise müssen die veröffentlichungspflichtigen Daten nur 10 Jahre vorgehalten werden, HmbGVBl. 2012, S.274.

¹⁷ DCAT-AP.de Pflegehandbuch, wie Anm. 10, S. 15.

DCAT ist im europäischen Kontext bereits als Standard etabliert. Sowohl das *European Data Portal* und das *European Union Open Data Portal* als auch Fachportale wie das *INSPIRE Geo Portal* und das *EUROSTAT Portal* arbeiten mit dem Vokabular.

In Deutschland ist die Verwendung von DCAT-AP.de nach dem zitierten Beschluss des IT-Planungsrates ab dem 01.01.2019 sowohl für die OpenData-Portale der Länder als auch für die zentralen Fachportale verpflichtend.

Entsprechend nimmt das nationale Datenportal GovData-Portal seit Jahresanfang sowohl von sämtlichen OpenData-Portalen der Länder als auch von den länderübergreifenden Fachdatenportalen wie z.B. *GeoPortal.de* und Destatis (Statistikportal) nur noch Daten im neuen Schema entgegen. Noch nicht vollständig realisiert werden konnte die Umstellung bei allen den Landesportalen zuliefernden Stellen, da hier, z.B. gegenüber den zuliefernden Kommunen, die rechtliche Möglichkeit fehlt, die Datenbereitsteller entsprechend zu verpflichten.

Zielsetzung DCAT-AP auf europäischer Ebene

Die Auflösung der Abkürzung AP (*Application Profile for Data Portals in Europe*) verdeutlicht die Zielsetzung der Ableitung. DCAT-AP wird vom Europäischen Datenportal verwendet sowie von weiteren Portalen aus 36 europäischen Ländern (Stand März 2019), die Daten an das Europaortal liefern.¹⁸ Nationale und anwendungsspezifische Anpassungen ermöglichen die Etablierung national verpflichtender Standards und die Einbeziehung besonderer Anwendungsfälle. So nutzen z.B. das INSPIRE-GeoPortal und das EUROSTAT Portal die spezifischen Ableitungen GeoDCAT-AP¹⁹ und StatDCAP-AP²⁰, um die Interoperabilität mit anderen offenen Verwaltungsdaten zu gewährleisten.

DCAT nutzt semantische Technologien zur Beschreibung der Daten und weitere Techniken aus dem Bereich des Linked Open Data. DCAT nutzt, fördert und verbreitet diese Techniken. Government Data bezeichnet im World Wide Web frei verfügbare Daten, die per Uniform Resource Identifier (URI) identifiziert sind und darüber direkt per HTTP abgerufen werden können. Ebenfalls per URI wird auf weitere, thematisch verwandte Datensätze verwiesen. Zur Beschreibung der Ressourcen, Kodierung und Verlinkung der Daten wird dabei RDF (Resource Description Framework) genutzt. So beschriebene Daten können dann mit der Semantic-Web-Abfragesprache SPARQL abgefragt werden, die in der Lage ist, RDF-Datensätze zu verlinken.²¹

Dabei stehen die folgenden Ziele im Vordergrund:

- Die Strukturen von DCAT unterstützen optimal Auffindbarkeit, Wiederverwendbarkeit, Verarbeitbarkeit von Datensätzen, die aus verschiedenen Systemen zusammengeführt werden.

18 Siehe URL: <https://learn.opengeoedu.de/opendata/vorlesung/open-government-data/dcat-ap> [Abruf 15. 5. 2019].

19 Siehe URL: <https://joinup.ec.europa.eu/solution/geodcat-application-profile-data-portals-europe> [Abruf 15. 5. 2019].

20 SieheURL: <https://joinup.ec.europa.eu/solution/statdcat-application-profile-data-portals-europe> [Abruf 15. 5. 2019].

21 European Commission (Hrsg.), OPEN DATA SUPPORT;Einführung in Linked Data, Slide 6 ff.

- Datenabschöpfung und Datennutzung über verschiedene Plattformen ist ohne Sprachbarrieren und mit nur einem Übernahmesystem möglich.
- DCAT eignet sich nicht nur zur differenzierten Darstellung einzelner Datensätze und ihrer Ressourcen (in DCAT Distributionen genannt), sondern auch zur Beschreibung der Kataloge von Datenkollektionen, wobei die Objekte jeweils mehrfach ineinander verschachtelt sein können.

Die Einhaltung des Standards gewährt Interoperabilität über die verschiedenen existierenden Portalebene(n) (von der kommunalen zur europäischen) und erlaubt eine erleichterte Auffindbarkeit der Datensätze, auch durch Suchmaschinen, auch über Landesgrenzen und fachliche Grenzen hinweg. So wird es für die verschiedenen Regierungsstellen leichter, Informationen auszutauschen und interoperable (Fach-)Systeme zu entwickeln.

Besonderheiten der nationalen Spezifikation DCAT-AP.de

Nationale Besonderheiten werden in der deutschen Ausprägung berücksichtigt, wovon hier die Wichtigsten aufgezeigt werden:

- Die nationale deutsche Spezifikation DCAT-AP.de macht verschiedene optionale Felder aus DACAT-AP zu Pflichtfeldern.
- Da eine notwendige Domänenneutralität von AP z.B. Angaben zu geographischen Einheiten beschränkt, füllen die nationalen Ableitungen diese Lücke. Die Geopolitische Definition durch Kennzeichnung der Verwaltungsebene Bund, Bundesland, Kreis, Kommune ist nur in Deutschland von Interesse und sieht verschiedene Varianten für den Raumbezug wie Regionalschlüssel vor.²²
- Die Zuordnung zur Pflichtklasse Kategorie mit deutschsprachigen Begrifflichkeiten erfolgt über die Struktur `dcat:theme` mit dem Vokabular `data:thema`, das zusammen mit den anderen DCAT-Vokabularen auf einer EU-Plattform gepflegt wird. Die eher kryptischen Schlüsselwörter EDUC (für Education) und JUST werden dann im nationalen Kontext zu *Bildung, Kultur und Sport bzw. Justiz, Rechtssystem und öffentliche Sicherheit*, was zudem die Sachverhalte exakter beschreibt.²³
- Angaben zu Lizenzdokumenten sind verpflichtend, sie referenzieren auf eine im GovData-Portal zu pflegende Liste.
- Deutschen rechtlichen Besonderheiten folgt die Hinzufügung zusätzlicher Rollen wie Urheber, Verwalter, Ansprechpartner und Autor.
- Um die Verbreitung nur valider Daten sicherzustellen, schreibt die deutsche Spezifikation zudem eine eigene verpflichtende Liste von unterstützten Hash-Algorithmen zur Erzeugung der Prüfsumme vor - eine Anforderung, die nach dem allgemeinen Anwendungsprofil auf freiwilliger Basis mitgeliefert werden kann.

Gliederung in Klassen

.....
 22 Soll ein Standard umfassend genutzt werden, kann auf Berücksichtigung nationaler Besonderheit nicht verzichtet werden. Die Umsetzung von Inspire zeigt, je genauer ein Datensatz beschrieben ist, desto wahrscheinlicher ist dessen Weiternutzung.

23 Vgl. Konventionenhandbuch, wie Anm. 11, S. 21.

Unterschieden nach den zu beschreibenden Inhalten teilt sich DCAT-AP.de in Klassen. Förderlich für Akzeptanz und Verbreitung ist die Vorgabe, dass nur 5 Klassen verpflichtend zu nutzen sind, 4 weitere werden empfohlen, und weitere 18 stehen für umfangreichere Beschreibungen zur Verfügung.²⁴

Die wichtigsten Klassen sowie ihr URI-Vokabular und Referenzen auf die jeweiligen Wertebereiche zeigt die folgende Tabelle:

Hauptklassen DCAT-AP.de	URI	Referenz/ Wertebereich
Verantwortliche Stelle	foaf:Agent	http://xmlns.com/foaf/spec/#term_Agent ; https://www.w3.org/TR/vocab-org/ ; foaf:Agent
Katalog	dcat:Catalog	http://www.w3.org/TR/2013/WD-vocabdcat-20130312/#classcatalog
Datenstruktur	dcat:Dataset	https://www.w3.org/TR/vocab-dcat/
Ressource	rdfs:Resource	http://www.w3.org/TR/rdfschema/#ch_resource
Kategorie	skos:Concept	dcat:theme, subproperty of dct:subject
Kategorien-Schema	skos:ConceptScheme	dcat:themeTaxonomy
Distribution	dcat:Distribution	http://www.w3.org/TR/2013/WD-vocabdcat-20130312/#classdistribution / dcat:Distribution
Lizenzdokument	dct:LicenseDocument	http://dublincore.org/documents/2012/06/14/dcmterms/?v=terms#LicenseDocument

Beispiele

Der im Transparenzportal Hamburg publizierte Datensatz mit dem Titel (dct:title) *Schüler- und Klassenzahlen an Hamburger allgemeinbildenden Schulen* (25.07.2017) wird im Transparenzportal Hamburg publiziert. Zusammen mit Tausenden anderen Ressourcen wird er mit einer DCAT-AP.de-Beschreibung an das GovData-Portal weitergereicht. Er führt als Beschreibungsdaten (dcat:description) den Inhalt *„Die Schüler- und Klassenzahlen an Hamburger Schulen werden jeweils zu Beginn des Schuljahres im Herbst erhoben. Die Zahlen - aggregiert nach Schulform, Schul- oder Jahrgangsstufe - können pro Bezirk, Stadtteil oder Schule ausgegeben werden. Sie können in Form einer Excel- oder pdf-Datei heruntergeladen werden.“* Er wird mit DCAT-AP.de im XML-Format mit Metadaten beschrieben (Auszug):

.....
24 Vgl. Spezifikation, wie Anm. 9, S. 14 ff.

```

<?xml version="1.0" encoding="UTF-8"?>
- <rdf:RDF xmlns:dcatde="http://dcat-ap.de/def/dcatde/1.0.1/" xmlns:vcard="http://www.w3.org/2006/vcard/ns#" xmlns:schema="http://schema.org/"
  xmlns:dct="http://purl.org/dc/terms/" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:spdx="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://www.w3.org/ns/foaf#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:spdx="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:locn="http://www.w3.org/ns/locn#" xmlns:spdx="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
- <dcat:Dataset rdf:about="https://ckan.govdata.de/dataset/a82b3fae-bfaa-404d-977b-1da9c965f780">
  - <dcat:distribution>
    - <dcat:Distribution rdf:about="https://ckan.govdata.de/dataset/a82b3fae-bfaa-404d-977b-1da9c965f780/resource/4d3ea43a-5cb9-4a99-b1fb-1434fe64599f">
      <dct:format>HTML</dct:format>
      <dct:license rdf:resource="http://dcat-ap.de/def/licenses/dl-by-de/2.0/">
      <dct:title>Link zum Downloadportal Schulstatistisches Informationssystem</dct:title>
      <dcat:byteSize rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">16356.0</dcat:byteSize>
      <dct:language>http://publications.europa.eu/resource/authority/language/DEU</dct:language>
      <dcat:accessURL rdf:resource="https://gateway.hamburg.de/HamburgGateway/Service/Entry/18/">
      - <spdx:checksum>
        - <spdx:Checksum rdf:nodeID="Ncdbc0d090a5f40999833b015453ece46">
          <spdx:checksumvalue
            rdf:datatype="http://www.w3.org/2001/XMLSchema#hexBinary">52fe1caf4f9b718edbcc36b3df37c38eeebadba400d605305b3d00ae784fa60</spdx:checksumvalue>
          </spdx:Checksum>
        </spdx:checksum>
      <dcatde:licenseAttributionByText>Freie und Hansestadt Hamburg, Behörde für Schule und Berufsbildung, [Jahr]</dcatde:licenseAttributionByText>
    </dcat:Distribution>
  </dcat:distribution>
  <dcat:keyword>Schuljahresehebung</dcat:keyword>
  - <dcatde:maintainer>
    - <foaf:Organization rdf:nodeID="Nb4b015acaa3347fe83d109ee7cd398f">
      <foaf:name>XXXXX, AXXXXXX, Herr Dr.</foaf:name>
    </foaf:Organization>
  </dcatde:maintainer>
  - <dct:creator>
    - <foaf:Organization rdf:nodeID="N96393bb5ef444c1dbd50dfa604201e12">
      <foaf:name>Behörde für Schule und Berufsbildung, Amt für Verwaltung, Referat V 12 (Datenmanagement)</foaf:name>
    </foaf:Organization>
  </dct:creator>
  - <dct:spatial>
    - <dct:Location rdf:nodeID="N5973b63a889f4109a3571dbdf29faaf2">
      <locn:geometry rdf:datatype="https://www.iana.org/assignments/media-types/application/vnd.geo+json">{"type": "polygon", "coordinates":
        [[[[10.326304, 53.394985], [10.326304, 53.964153], [8.420551, 53.964153], [8.420551, 53.394985], [10.326304, 53.394985]]]]
      </locn:geometry>
    </dct:Location>
  </dct:spatial>
  <dct:title>Schüler- und Klassenzahlen an Hamburger allgemeinbildenden Schulen</dct:title>
  - <dcat:distribution>
    - <dcat:Distribution rdf:about="https://ckan.govdata.de/dataset/a82b3fae-bfaa-404d-977b-1da9c965f780/resource/cd69f465-ced3-4fc7-8b67-ef1445579ef5">
      <dcat:byteSize rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">48907.0</dcat:byteSize>
      <dct:language>http://publications.europa.eu/resource/authority/language/DEU</dct:language>
      <dct:format>HTML</dct:format>
      <dcatde:licenseAttributionByText>Freie und Hansestadt Hamburg, Behörde für Schule und Berufsbildung, [Jahr]</dcatde:licenseAttributionByText>
      <dct:title>Metadatenbeschreibung aus dem MetaVer</dct:title>
      <dct:license rdf:resource="http://dcat-ap.de/def/licenses/dl-by-de/2.0/">

```

Abb. 2: Beschreibung Beispieldatensatz Schüler und Klassen in DCAT-AP.de (Auszug)

Die Betrachtung verschiedener Kategorien von weiteren Beispieldatensätzen aus dem Transparenzportal Hamburg verdeutlicht die weitreichende Integration vorhandener Metadatenysteme wie Dublin Core Metadata Terms (dct) und anderer Vokabulare.

Räumliche Abdeckung

Ein räumlicher Bereich oder ein bezeichneter Ort definiert den Raumbezug des Datensatzes. Er kann durch ein kontrolliertes Vokabular oder mit geographischen Koordinaten repräsentiert werden und bietet wertvolle Zusatzinformationen für die Eingrenzung bei Recherchen und für räumliche Darstellungen:

```

<dct:spatial>
  <dct:Location rdf:nodeID="N5973b63a889f4109a3571dbdf29faaf2">
    <locn:geometry rdf:datatype="https://www.iana.org/assignments/media-types/application/
      vnd.geo+json">{"type": "polygon", "coordinates": [[[[10.326304, 53.394985], [10.326304,
      53.964153], [8.420551, 53.964153], [8.420551, 53.394985], [10.326304, 53.394985]]]]
    </locn:geometry>
  </dct:Location>
</dct:spatial>

```

Prüfsumme

In DCAT-AP.de gehört die Checksumme zu den unverzichtbaren Elementen. Der Wert ermöglicht die Verifizierung des Datensatzes, auch wenn er bereits über zahlreiche Stationen weitergereicht wurde. Diese Klasse ermöglicht es, die Ergebnisse einer Vielzahl von Prüfsummen- und Krypto-Algorithmen zu repräsentieren.

```

<dc:Distribution rdf:about="https://ckan.govdata.de/dataset/006edaff-d869-4cbc-91c3-
d8599a548067/resource/b5c3f3be-2241-4b8c-a3b5-a00587469706">
  <dc:language>http://publications.europa.eu/resource/authority/language/DEU</dc:language>
  <dc:format>JPEG</dc:format>
  <dc:license rdf:resource="http://dcat-ap.de/def/licenses/dl-by-de/2.0"/>
  <dc:licenseAttributionByText>Freie und Hansestadt Hamburg, Landesbetrieb Geoinformation und
  Vermessung </dc:licenseAttributionByText>
  <dc:title>Download DOP 20 , Sommer 2013, (JPG)</dc:title>
  <dc:license rdf:resource="http://dcat-ap.de/def/licenses/dl-by-de/2.0"/>
  <dc:accessURL rdf:resource=
  "http://daten-hamburg.de/geographie_geologie_geobasisdaten/digitale_orthophotos/
  DOP20_belaubt/DOP20_HH_sommerbefliegung_2013.zip"/>
  <spdx:checksum>
    <spdx:Checksum rdf:nodeID="N2ba7aff1b7d644b7a7a9fe5873703a56">
      <spdx:checksumValue
      rdf:datatype="http://www.w3.org/2001/XMLSchema#hexBinary">4309234332</spdx:checksumValue>
    </spdx:Checksum>
  </spdx:checksum>
</dc:Distribution>

```

Kontakt, Herausgeber

Bei Werten der Felder Herausgeber (Datenbereitsteller), Kontakt, Identifier etc. dürfen aus Gründen des Datenschutzes keine personenbezogenen Daten eingetragen werden. Hier tauchen dann auch nicht mehr namentlich genannte Personen auf sondern organisatorische Einheiten.

```

<dc:title>Beteiligungsbericht 2015</dc:title>
<dc:contactPoint>
  <vcard:Organization rdf:nodeID="Nd3604e9d1cd54b9fa20b524ca4d0ed00">
    <vcard:fn>Fachliche Leitstelle Transparenzportal</vcard:fn>
  </vcard:Organization>
</dc:contactPoint>
<dc:publisher>
  <foaf:Organization rdf:about="https://ckan.govdata.de/organization/1b414196-261e-4fbf-8b4b-
30756dcd8ea1">
    <foaf:name>Transparenzportal Hamburg</foaf:name>
  </foaf:Organization>
</dc:publisher>

```

Zusammenfassend kann man sagen, dass Datensätze, die mit DCAT-AP.de beschrieben werden, neben den für Archive zwingend notwendigen Metadaten wie Titel und Entstehungszeitraum eine Reihe relevanter Metadaten liefern. Diese werden aktuell zwar nicht von EAD oder archivischen Verzeichnungsstandards gefordert, jedoch können bei der Übernahme in ein digitales Archiv und bei einer späteren Nutzung dieser Daten einen deutlichen Mehrwert für Bereitstellung und Nutzende haben.

Beispiel TILDA (Schweden) – Metadatenverarbeitung von Forschungsdaten

In den traditionell transparenteren Administrationen der skandinavischen Länder werden deutlich mehr Informationen der staatlichen Verwaltung, aber auch Daten von Firmen und Privatleuten, publiziert und intensiv genutzt. Entsprechend ist DCAT-AP schon seit Jahren etabliert. Bei diesem Erfah-

runghintergrund ist es naheliegend, archivische Projekte einzurichten, die DCAT-AP als universelle Schnittstellen für die digitale Archivierung zu nutzen versuchen.

In Schweden entwickelt die *Universität für Agrarwissenschaften (SLU)* das System TILDA zur Verbreitung und OAIS-konformen Archivierung von Forschungs- und Umweltdaten.²⁵ Der Schwerpunkt liegt auf der Erstellung einheitlicher und klarer Routinen und Methoden für das Datenmanagement, einem System zur Unterstützung der Registrierung von Metadaten und der elektronischen Archivierung. Daneben sollen - ebenfalls auf den Metadaten von DCAT-AP basierend - die Unterlagen aus dem digitalen Archiv bereitgestellt werden (OAIS-Access) sowie die Möglichkeit integriert werden, Datensätze mit Publikationen zu verknüpfen.

Das Modul TILDA-Ingest realisiert dabei die Übernahme von Daten aus verschiedenen Datenkollektionen mit unterschiedlichen Metadaten. Um semantische und technische Interoperabilität zu erreichen, werden dabei unterschiedliche Metadatenstandards berücksichtigt. TILDA generiert transformiert SIP-Datenpakete auf der Basis von DCAT, aber auch mit METS, Geodata und anderen beschriebenen Metadatenstandards und den dazugehörigen Ressourcen.

Ein weiteres wichtiges Ziel dieses digitalen Archivsystems ist die Kommunikation mit verschiedenen Fachportalen wie OpenAIRE, EUDAT, dem Geoportal INSPIRE und der künftigen schwedischen nationalen Forschungsdateninfrastruktur, mit der Option, Datensätze mit Publikationen zu verknüpfen. Da bei fortschreitender Etablierung von DCAT-AP davon auszugehen ist, dass künftige Archivnutzer bei der Bereitstellung von in digitalen Archiven verwahrten Rohdaten erwarten, dass diese bei Vorlage auch in DCAT-AP.de beschrieben sind, sollten sich Archive auch mit Blick auf das OAIS-Modul Access mit dem neuen Metadatenstandard für offene Verwaltungsdaten auseinandersetzen.

Schlussstatement

DCAT-AP.de stellt Instrumente bereit, um offene Verwaltungsdaten in digitale Archive zu übernehmen. Die auch unter dem Begriff der offenen Verwaltungsdaten zu subsumierenden behördlich publizierten Registraturdaten sind dafür ungeeignet, da es sich dabei in der Regel allenfalls um Ergebnisdokumente handelt. Dagegen bietet sich die große Anzahl von Datensammlungen mit Rohdaten aus den Bereichen Statistik, Messdaten, Geodaten und Geofachdaten an. Im Einzelfall wird zunächst zu prüfen sein, ob die mitgelieferten Metadaten für eine Übernahme ins Archiv ausreichen. Beispielsweise sind die im Transparenzportal Hamburg publizierten Daten regelhaft mit einer Fülle von Metadaten beschrieben, die im Normalfall über die Notwendigkeiten der Digitalen Archivierung hinausreichen. Anfang des Jahres 2019 wurden vom Landesamt

25 Vgl. https://snd.gu.se/sites/default/files/legacy/Tilda_publicering%20och%20e-arkivering%20av%20forskningsdata_0.pdf und <https://www.opensciencefair.eu/posters/tilda-a-solution-at-slu-sweden-for-e-archiving-long-term-preservation-and-publishing-of-research-and-environmental-data> [Abruf 4. 10. 2019].

für Geoinformation und Vermessung mit einem eigenen Ingest-Prozess aus den Speichersystemen der Behörde Orthofotos der Bildflüge des Jahres 2001 in das digitale Archiv des Staatsarchivs übernommen. Auf Basis der DCAT-AP.de-Metadaten des Transparenzportals Hamburg hätte dies ohne Informationsverlust und mit der Gewinnung zusätzlicher relevanter Optionen ebenso realisiert werden können.²⁶

Anhand von verschiedenen Beispielen wäre zu prüfen, ob es möglich ist, einen Ingest für das digitale Archiv zu realisieren, der die Übernahme verschiedenster mit DCAT-AP.de beschriebener Datensammlungen realisiert. Im Idealfall entstünde eine universelle Schnittstelle, die in der Lage wäre, aus unterschiedlichsten, mit DCAT-AP.de beschriebenen Datensammlungen, OAIS-konforme SIPs zu formieren und ins digitale Archiv zu transportieren. So ließe sich die Übernahme von Datensammlungen unterschiedlichster Provenienz standardisieren und deutlich vereinfachen.

Summary

German federal, state and local authorities have a considerable amount of data collections which are to be made available to the public. DCAT-AP.de is the German specification of the European Data Catalogue Vocabulary Application Profile (DCAT-AP). It is the obligatory European metadata standard for Open Government Data which is used by the European Data Portal and 36 European countries. Taking into account the rules for Linked Open Data DCAT is designed to link and standardise data catalogues from various sources. The standard further facilitates interoperability between different national and international portal levels. In July 2018 DCAT-AP.de was defined standard for Open Government Data by the German IT Planning Council.

DCAT-AP.de provides the instruments to integrate data collections containing raw data from the sectors statistics, measuring data, such as basic and customized geodata into digital archives. Ingest tools that can handle DCAT-based data collections are already in the works. Ideally an europe-wide interface which is able to form OAIS compliant SIPs using different DCAT-AP coded databases will become reality soon.

.....
 26 Archivrechtliche und organisatorisch gäbe es bei dieser Vorgehensweise verschiedene Klärungsbedarfe. So wäre z.B. die Umwandlung der Unterlagen in Archivgut zu regeln sowie mit der abgebenden Behörde über den Rechtsstatus der Daten eine Verabredung zu treffen.

ERHALTUNGS-
STRATEGIEN
UND
ERHALTUNGS-
MASSNAHMEN

ALLES BLOCKCHAIN ODER WAS?

Nutzen und Grenzen der Distributed-Ledger-Technologie in der digitalen Archivierung

Björn Steffenhagen

Einführung

Blockchain ist ein Hype. Es vergeht kaum ein Tag, in dem in den Medien nicht von einem neuen Blockchain-Projekt berichtet wird. Mittels eines dezentralen Netzwerks soll ein authentischer und integrierter Austausch von Informationen stattfinden, durch den Einsatz kryptografischer Verfahren ist die unbemerkte Manipulation ausgeschlossen. So sind verschiedenste Einsatzszenarien denkbar: über die bekannten Kryptowährungen wie Bitcoin, das digitale Grundbuchamt, der Finanzindustrie und auch die Implementierung innerhalb ganzer Industrie-Sektoren. Viele neigen daher dazu, von einer Revolution der internetbasierten Kommunikation zu sprechen¹ – und wecken damit hohe Erwartungen. Andere Stimmen kritisieren hingegen diesen Ansatz und verweisen auf fehlende Anwendungsfälle. So gäbe es lediglich spezifische Anwendungsfälle, in denen eine Blockchain, im Gegensatz zu einer zentralen Datenbanklösung, als sinnvollere Lösung erscheint.²

Im vorliegenden Artikel soll daher abseits des Hypes den Fragen nachgegangen und allgemeinverständlich erläutert werden, was Blockchain beziehungsweise die Distributed Ledger-Technologie eigentlich ist und welchen Grundsätzen sie folgt. Hierzu soll der Ansatz von Blockchain kurz in einen Gesamtkontext gestellt und eine Aussage zum Verhältnis der Blockchain zu unseren bisherigen Standards getroffen werden. Das Beispiel der National Archives UK zeigt auf, wie eine Gedächtniseinrichtung die praktische Implementierung einer Blockchain vollzogen hat. Im Weiteren folgen eine mögliche Architektur einer Blockchain für die Umsetzung in der digitalen Langzeitarchivierung, basierend auf der aktuellen Forschung, den Chancen und Risiken der Technologie sowie ein Fazit.

Grundlagen

Eine Blockchain oder auch Distributed-Ledger ist ein verteiltes Protokollregister mehrerer gleichberechtigter Teilnehmer, die untereinander Daten austau-

-
- 1 Vgl. bspw. Tapscott, Alex; Tapscott, Don: Die Blockchain-Revolution. Wie die Technologie hinter Bitcoin nicht nur das Finanzsystem, sondern die ganze Welt verändert, Wiesbaden 2016.
 - 2 Vgl. Wieler, Herbert: Blockchain. Was ist der Unterschied zu einer herkömmlichen Datenbank? o. O. 2018. URL: <https://www.infopoint-security.de/Blockchain-was-ist-der-unterschied-zu-einer-herkoemmlichen-datenbank/a15631/> [Abruf 4. 10. 2019].

schen. Die durchgeführten Transaktionen werden bei sämtlichen Teilnehmern vollständig und redundant vorgehalten. Die Blockchain selbst dient dazu, die Integrität der Daten bzw. Dateien sicherzustellen sowie die Transaktionen nachweisbar zu halten. Durch das Zusammenfassen von Transaktionen und einer chronologischen Verkettung mittels Hashwerten entsteht ein robustes und fälschungssicheres Protokoll.³ Jeder Nutzer ist gleichzeitig auch ein Betreiber des Systems, ein solches dezentrales Netzwerk benötigt daher bei einer vollständigen Umsetzung keinen zentralen Intermediär mehr.⁴ Aus technischer Sichtweise kann ein solcher Intermediär in einem Netzwerk z. B. eine zentrale Datenbankanwendung darstellen. Einer breiteren Öffentlichkeit wurde das Grundprinzip einer Blockchain in 2008 bekannt, als Satoshi Nakamoto⁵ ein White Paper zur Architektur der Kryptowährung Bitcoin veröffentlichte.⁶

Wenn jedoch keine zentrale Instanz mehr existiert, die kontrollieren und bestimmen kann, welche Inhalte valide sind, dann muss auch hier der dezentrale Ansatz verfolgt werden. Transaktionen werden in einer Blockchain erst ausgeführt, wenn ein Konsens unter der Mehrheit der Teilnehmer erzeugt wurde – die Mehrheit bestimmt somit, welche Inhalte als valide gelten. Es gibt verschiedene Mechanismen, wie dieser Konsens erzeugt wird. Zwei der bekanntesten sind der Proof-of-Work und der Proof-of-Stake. Der Proof-of-Work-Mechanismus geht grundsätzlich davon aus, dass eine nachweisbare Arbeit stattfinden muss, um einen bestätigten Eintrag in das digitale Register zu erhalten. Als Aufgabe gilt hier die Lösung kryptografischer Rätsel – In der Bitcoin-Blockchain auch unter dem Begriff „Mining“ bekannt. Dieses Verfahren ist äußerst aufwendig und benötigt Milliarden an Rechenoperationen. Die Validierung hingegen ist durch die Hashwertberechnung des berechneten Ergebnisses und dessen Abgleich sehr simpel.⁷ Dieses Verfahren gilt als sicher und robust, gleichzeitig ist in den meisten Fällen ein Honorierungssystem vorgesehen, um die Beteiligung an diesem Konsensmechanismus möglichst attraktiv zu gestalten. Das hieraus resultierende Problem des intensiven Hardware-Einsatzes und hohen Stromverbrauchs möchte der Proof-of-Stake umgehen. Hier werden zu validierende Objekte je nach Besitz unter den Teilnehmern verteilt. Je mehr Daten ein Teilnehmer in dieser Variante vorhält, desto größer ist die Wahrscheinlichkeit, dass dieser auch andere Daten validiert. Prinzipiell wird dabei von einem Grundbedürfnis zum Systemerhalt der Nutznießer ausgegangen. Dieser Besitz kann dabei ein bestimmtes Guthaben einer Währung sein. Eine entsprechende Regel kann aber auch der Besitz gültiger Anmeldedaten und einer verifizierten Identitätsfeststellung darstellen. Die Teilnehmer an ei-

.....
 3 Vgl. Glatz, Florian; Holthusen, Jannis; Kufeld, Simon: Vorstellung der Blockchain-Technologie. o. O. 2016, S. 2. URL: <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Vorstellung%20der%20Blockchain-Technologie.pdf> [Abruf 4. 10. 2019].

4 Vgl. Wieler, Herbert: Blockchain.

5 Ein bis heute nicht aufgedecktes Pseudonym für eine Person-, bzw. Personengruppe.

6 Nakamoto, Satoshi: Bitcoin. A Peer-to-Peer Electronic Cash System. o. O. 2008. URL: <https://bitcoin.org/bitcoin.pdf> [Abruf 4. 10. 2019].

7 Vgl. Roth, Matthias: Wie funktioniert eine Blockchain? o. O. 2017. URL: <https://www.adeso.de/de/news/blog/wie-funktioniert-eine-Blockchain.jsp> [Abruf 4. 10. 2019].

ner Blockchain vertrauen sich grundsätzlich nicht gegenseitig und müssen sich auch nicht persönlich kennen. Vielmehr werden als vertrauensbildende Maßnahmen u. a. digitale Signaturen, eine Public-Key-Infrastruktur sowie die angesprochenen Hashwerte⁸ genutzt.

Um die verschiedenen Prozesse in einer Blockchain zu automatisieren, werden sogenannte smart contracts genutzt. Smart contracts sind Programmcodes innerhalb der Blockchain, die einige Eigenschaften dieser geerbt haben. Sie sind ebenso unveränderlich und werden bei jedem Teilnehmer vorgehalten. Nach definierten Regeln sind hier Vertragsdetails hinterlegt, welche die Kommunikation zwischen den menschlichen und maschinellen Teilnehmern automatisieren.⁹ Ein simples Beispiel wäre der Kauf eines Hauses in einer Blockchain ohne die Bestätigung eines Notars, wobei durch smart contracts nach Zahlungseingang die Eigentumsumschreibung erfolgt wie auch der automatisierte Zugang zum Haus ermöglicht wird. Ähnliche Szenarien sind für die digitale Archivierung denkbar: Der Nutzer eines Archivs stellt einen Antrag auf Schutzfristenverkürzung. Nach Prüfung und Genehmigung gibt die Behörde das Archivgut zur Einsicht frei. Durch einen smart contract wird das digitale Archivgut aus dem Archivspeicher geholt, geprüft, ob es authentisch und integer ist, eventuelle Restriktionen dem Nutzer mitgeteilt und technisch durchgesetzt.

Architektur

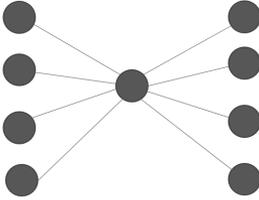
Der traditionelle Ansatz für Transaktionen geht von mehreren Teilnehmern aus, die über eine zentrale Stelle, einem Intermediär, Informationen, materielle oder immaterielle Güter austauschen. Ein direkter Kontakt zwischen den Teilnehmern besteht nicht. Auch ist diese Zentrale alleiniger Besitzer aller nötigen Informationen in diesem Netzwerk. Der Ansatz einer Blockchain ist demgegenüber völlig konträr. In einem klassischen Ledger-Netzwerk ist jeder Teilnehmer gleichzeitig auch Besitzer sämtlicher Informationen. Die Server und Clients, auch Nodes genannt, laden jede jemals getätigte Transaktion in ihr System und sind im Weiteren für die Überprüfung aller weiteren Transaktionen zuständig.¹⁰

8 Im Blockchain-Umfeld werden i. d. R. die Algorithmen SHA-256 bzw. SHA-512 angewandt.

9 Vgl. Ebersbach, Jan: Blockchain-Anwendungen. Blockchain im praktischen Einsatz. In: iX. Magazin für professionelle Informationstechnik, Nr. 7 (2018), S. 50-54, hier: S. 51; Bundesverband IT-Sicherheit: TeleTrust-Positionspapier "Blockchain". Handreichung zum Umgang mit der Blockchain. Berlin 2017, S. 26. URL: https://www.teletrust.de/fileadmin/docs/publikationen/broschueren/Blockchain/2017_TeleTrust-Positionspapier_Blockchain__.pdf [Abruf 4. 10. 2019].

10 Vgl. Antonopoulos, Andreas: Bitcoin & Blockchain. Grundlagen und Programmierung. Die Blockchain verstehen, Anwendungen entwickeln. 2. Aufl. Heidelberg 2018, S. 174f.

zentrales Netzwerk



verteiltes, dezentrales Netzwerk

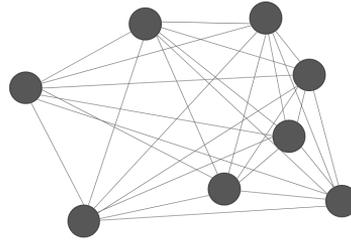


Abb. 1: Netzwerke¹¹

Die Blockchain steht für eine authentische, integre und nachweisbare Abwicklung von Transaktionen. Dieser Ansatz beruht wesentlich auf dem Aufbau der einzelnen Blöcke. Transaktionen werden grundsätzlich zu Blöcken zusammengefasst. Für jede Transaktion wird ein Hash generiert, ebenso für die Gesamtheit aller Transaktionen - der sogenannte Hash-Baum - und den Block selbst. Im folgenden Block wird der Hash referenziert, wodurch eine Verkettung entsteht – die Blockchain. Das Zusammenfügen erfolgt chronologisch, weshalb der erste Block, der Genesis-Block, dabei natürlich keinen Vorgänger aufweisen kann. Der zuletzt hinzugefügte Block kann auf keinen Nachfolger verweisen, was ein potenzielles Sicherheitsrisiko darstellt. Angreifer könnten versuchen kompromittierte Informationen als Nachfolger einzufügen. Um es Angreifern über den Netzwerkansatz und des aufwendigen Verfahrens zur Validierung faktisch unmöglich zu machen, invalide Transaktionen durchzuführen, greift an dieser Stelle der Konsensmechanismus. Ein möglicher Angriff auf ein Ledger-Netzwerk wäre effektiv nur dann möglich, wenn ein Angreifer 51% sämtlicher Nodes und damit den Konsensmechanismus kontrollieren könnte.¹²

stark vereinfacht

Genesis-Block

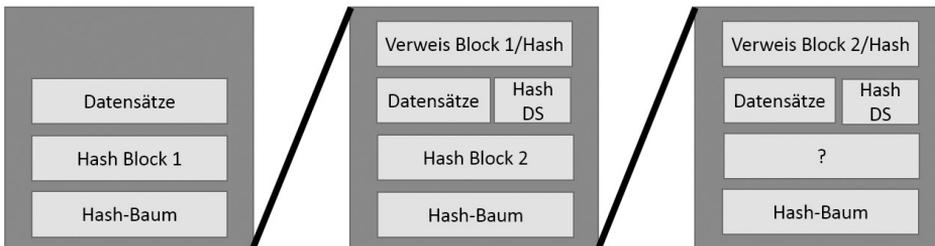


Abb. 2: Blockchain-Architektur¹³

11 Vgl. beispielhaft Eden. Jan: Die Geschichte des Internet. Köln o. J. URL: <https://eden.one/die-geschichte-des-internet> [Abruf 4. 10. 2019]

12 Vgl. Seeger, Jürgen: Von Bitcoin bis Ripple. Hyperledger-Implementierungen im Vergleich. In: iX. Magazin für professionelle Informationstechnik, 7 (2018), S. 44-48, hier S. 46.

13 Vgl. Burgwinkel, Daniel (Hg.): Blockchain Technology. Einführung für Business und IT-Manager. Berlin, Boston 2016, S. 6.

Je nach Art des Ledger können selbstverständlich unterschiedliche Informationen in einem Block enthalten sein, die meisten orientieren sich jedoch an zentralen Inhalten:

Identifizier
Difficulty (Schwierigkeitsgrad für den Proof-of-Work)
Nonce (Vorgabe für den Wert des Proof-of-Work)
Hash-Wert aktueller Block
Hash-Wert Vorgängerblock
Zeitstempel
Weitere Signaturen
Hash-Baum (Merkle) der durchgeführten Transaktionen
Transaktionsdaten

Abb. 3: *Beispiel eines Blocks*¹⁴

Wenn eine im Vorhinein definierte Menge an Transaktionen erreicht ist, wird ein Block – eine technische Zusammenfassung – aus diesen gebildet. Neben allgemeinen Informationen zum Block und den Primärdaten befinden sich hier auch Angaben zum Proof-of-Work, also unter welchen Bedingungen der Block zur Blockchain hinzugefügt wurde, um eine spätere Validierung zu ermöglichen. Neben seinem eigenen Hash enthält jeder Block auch den Hash seines Vorgänger-Blocks. Zeitstempel und digitale Signaturen dienen dem Nachweis authentischer Datenhaltung, ebenso wie der paarweise gebildete Hash-Baum¹⁵ der Primärdaten bzw. der Transaktionen.

Ausprägungen

- Nach Zweck
 - Kryptowährung
 - Transaktionsnachweis
- Nach Zugang
 - public Blockchain: Bitcoin
 - private Blockchain
- Nach Rechten/Besitz
 - Zugriffskontrolle
 - Identitätsnachweis
- Nach Betreiber

Abb. 4: *Ausprägungen*¹⁶

14 Vgl. Burgwinkel (Hg.): Blockchain Technology, S. 6; Antonopoulos, Andreas: Bitcoin & Blockchain., S. 203.

15 Auch Merkle-Tree genannt, nach dem Erfinder Ralph Merkle.

16 Vgl. Burgwinkel, Daniel: Blockchain Technology, S. 34f.; Keller-Marxer, Peter. Blockchains als

Wie dargestellt, sind nicht alle Inhalte eines Blocks verpflichtend. Dementsprechend gibt es auch verschiedene Möglichkeiten eine Blockchain zu gestalten. Die Kriterien, wie ein Ledger aufzubauen ist, können sich so daran ausrichten, welcher Zweck verfolgt wird. Eine Kryptowährung verlangt wesentlich höhere Sicherheitsanforderungen an die Architektur, als ein vergleichbares Ledger, welches zum einfachen Nachweis dient. Von Bedeutung ist weiter, wer auf das System Zugriff hat und welche Rechte die Teilnehmer besitzen sollen. Damit hängt auch die Frage zusammen, wer im Produktivbetrieb der eigentliche Betreiber, ggf. der Verantwortliche sein soll.¹⁷

Vor- und Nachteile

Aus der Betrachtung der Charakteristika der Blockchain-Technologie ergeben sich die spezifischen Vor- und Nachteile. Die Blockchain spielt ihre Vorteile in der Vertrauenswürdigkeit aus. Durch die redundante Datenhaltung kann stets ein authentischer Datenbestand durch unabhängige Drittparteien abgefragt werden. Gegenüber herkömmlichen zentralen Datenbanken ist die Manipulation einer Blockchainbasierten Infrastruktur wesentlich aufwendiger und kostenintensiver. Gleichzeitig können gewisse Prozesse beschleunigt, und Prozesskosten verringert werden, wenn Intermediäre wegfallen bzw. durch smart contracts ersetzt werden.¹⁸ Neue Herausforderungen lassen sich durch Technologien jedoch nicht vermeiden. Insbesondere public Blockchains, wie die Kryptowährungen, erfordern durch ihren Proof-of-Work einen hohen Ressourcenverbrauch und sind nur schwer skalierbar. Auch stellen sich für viele potentielle Anwender grundsätzliche und berechnete Fragen: Wofür sollen wir die Blockchain einsetzen? Welche Anwendungsfelder bedient sie? Was verbessert sie, dass nicht jetzt schon mit den derzeitigen technischen Mitteln erledigt werden kann?

Blockchain und Standards

Nach dem Einblick in den Ansatz von Blockchain soll an dieser Stelle ein kurzer Vergleich zu den eingesetzten Standards und Normen im Records Management und der digitalen Archivierung angestellt werden. Sämtliche Standards in der digitalen Archivierung entsprechen dem etablierten Ansatz zentraler Datenhaltung oder auch zentraler Vertrauensdienste. Von daher überrascht

.....
Herausforderung für das Records Management. In: Informationswissenschaft: Theorie, Methode und Praxis 5 (2018), S. 268-297, hier S. 285. URL: <https://doi.org/10.18755/iw.2018.30> [Abruf 4. 10. 2019].

17 Vgl. dazu auch Keller-Marxer, Peter: Blockchains als Herausforderung des Records Management, S. 270.

18 Vgl. beispielhaft Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V.: Online-Konsultation zur Erarbeitung der Blockchain-Strategie der Bundesregierung. Bitkom-Stellungnahme. Berlin 2019, S. 27. URL: https://www.bitkom.org/sites/default/files/2019-04/190408_stellungnahme_Blockchain-strategie-konsultation_online.pdf [Abruf 4. 10. 2019].

die Aussage nicht, dass ein offenes, zugangsfreies public Ledger – wie es der ursprüngliche Ansatz vorsah – nicht kompatibel mit den Standards und Verordnungen des Records Managements ist.¹⁹ Als ein wesentliches Beispiel sei der nicht erbringbare Identitätsnachweis durch Pseudonymisierungen genannt, welchen eIDAS jedoch vorsieht.²⁰ Jeder Teilnehmer einer Blockchain ist gleichzeitig auch Betreiber dieser, was vorgeschriebene und notwendige Fragen zur Zuständigkeit und Haftung im Störfall unmöglich macht.²¹

Auch wenn in der praktischen Umsetzung der digitalen Archivierung derzeit keine qualifizierten Signaturen oder Zeitstempel als rechtssichernde Verfahren eingesetzt werden, so gelten doch das OAIS-Modell oder die DIN 31644 als de-facto-Standards. Aus dem Blickwinkel von OAIS kommt deren Offenheit der Blockchain zugute. Zwar verweist OAIS ebenso auf die angesprochenen Punkte, schränkt die Nutzung durch weitere Kriterien jedoch nicht ein. Genauso ist die Möglichkeit der Migration veralteter Dateiformate durch eine strikte Nutzung des Repräsentationenmodells gegeben. Aus Sicht der Vertrauenswürdigkeit eines digitalen Archivs stellen insbesondere die Kriterien zur Verantwortung für den Erhalt, die technische Hoheit über das System und der Datenschutz Dritter Ausschlusskriterien für den Einsatz einer klassischen Blockchain dar.

Beispiel: Archangel der National Archives UK

Eines der ersten Projekte zur Implementierung einer Blockchain in einer Gedächtniseinrichtung setzte das Nationalarchiv Großbritanniens um. Archangel, ein mit rund 500.000 Pfund gefördertes Nationalprojekt, soll ermitteln, ob und inwieweit die Erhaltung der Integrität und der authentische Herkunftsnachweis eines digitalen Objekts in einem digitalen Archiv mit einer Blockchain umgesetzt werden kann. Das Projekt läuft seit Juni 2017 und wird im Juni 2019 beendet. Als Projektpartner fungieren Archivare, die im Forschungsdatenmanagement der Universität Surrey tätig sind.

.....
19 Vgl. Kampffmeyer: Konsistentes Löschen in der Block Chain. Hamburg 2019, S. 3. URL: https://www.project-consult.de/files/20190425_Konsistentes_Loeschen_in_der_Block_Chain_Kff.pdf [Abruf 4. 10. 2019].

20 Vgl. Verordnung (EU) Nr. 910/2014 des Europäischen Parlaments und des Rates über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG [eIDAS-VO]. Art. Brüssel 2014, S. 26. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32014R0910&from=DE> [Abruf 4. 10. 2019].

21 Vgl. Verordnung (EU) Nr. 910/2014 des Europäischen Parlaments und des Rates über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG [eIDAS-VO]. Art. 24, Abs. 1 & Abs. 2 c. Brüssel 2014, S. 26. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32014R0910&from=DE> [Abruf 4. 10. 2019].

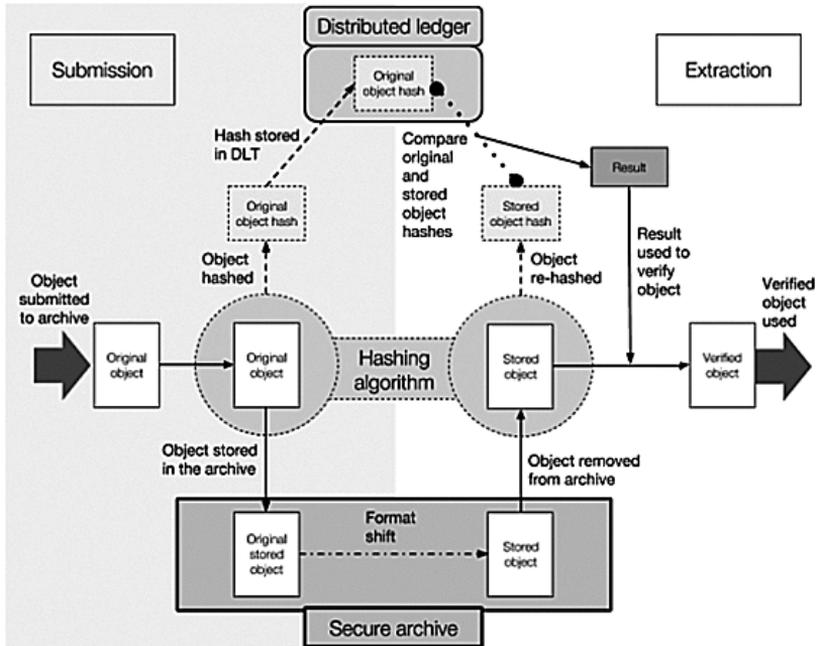


Abb. 5: Architektur Archangel²²

Die entsprechende Architektur sieht folgendermaßen aus: Bei der Übernahme eines digitalen Archivales wird ein SHA-256 Hash²³ generiert und zusammen mit Identifiern sowie beschreibenden Metadaten in die Blockchain überführt. Das Archivgut gelangt über den Ingest weiterhin regulär und OAIS-konform in das digitale Archiv der National Archives. Beim Herausholen einer Kopie aus dem Archivspeicher im Rahmen des Access wird erneut der Hash generiert und gegen den in der Blockchain befindlichen geprüft. Bei Übereinstimmung gilt das Objekt als authentisch und integer und kann dem Nutzer ausgegeben werden. Eine Formatkonvertierung gilt hier als neuer Ingest, bei der der Hash neu erzeugt und in der Blockchain hinterlegt wird. Im Weiteren ist es angedacht, neue Kooperationspartner zu akquirieren und bei diesen ebenfalls Ledger zu installieren. Der Zugriff auf die Archivspeicher sowie dem Ledger bleibt dabei in der Verantwortung der National Archives. Durch smart contracts automatisiert, werden die Transaktionsdaten des Ingest zu Blöcken gefasst und mit einem Proof-of-Work bestätigt.²⁴ Neben der Implementierung und Erprobung der Technik und Prozessabläufe sind für beide Projektpartner vor allem die "weichen" Faktoren als erste Projektergebnisse erwähnenswert: Das neue Verfahren wurde im Labortest positiv angenommen. Gerade die dezentrale Lösung des Ledger schafft die Notwendigkeit miteinander zu arbeiten, ggf. auch über Sparten- und Ländergrenzen hinweg und legt so

22 Vgl. Collomosse, John et. al.: ARCHANGEL: Trusted Archives of Digital Public Documents. In: ACM Document Engineering 2018. o. O. 2018, S. 2. URL: <https://arxiv.org/abs/1804.08342> [Abruf 4. 10. 2019].

23 Als Alternative werden auch formatspezifische, neuronale Netzwerke angegeben, die individuelle Charakteristika einer Datei verarbeiten, vgl. ebd.

24 Vgl. ebd., S. 2f.

auch den Grundstein für weitere Kooperationen. Als weiteren Punkt wird die Vertrauenswürdigkeit genannt, die sich von der einzelnen Einrichtung hin zu einem Vertrauen in das Gesamtsystem verlagert, was in Zeiten von Fake-News und Vertrauensverlust in öffentliche Einrichtungen – so die Ersteller des Artikels – keine zu unterschätzenden Faktoren darstellen.²⁵

Ansatz einer möglichen Architektur einer Blockchain in der digitalen Archivierung

Wie sieht also der Rahmen aus, in dem eine Blockchain produktiv in der digitalen Archivierung eingesetzt werden kann? Eine mögliche Blockchain-Architektur muss innerhalb eines Archivverbundes stattfinden, um eine Validierung außerhalb der eigenen Zuständigkeit zu ermöglichen. Eine lediglich hausintern eingesetzte Lösung erscheint dagegen in sich obsolet, da Authentizität und Integrität, wenn möglich, von unabhängigen, qualifizierten Anbietern geprüft werden muss. Statt der Primärdaten sollte lediglich ein Grundset an Metadaten und der Hash in der Blockchain hinterlegt werden, um eine Identifikation und Referenzierung der Blöcke zu gewährleisten. Zur Validierung sollte ein ressourcenschonender Proof-of-Stake eingesetzt werden. Als private und zugangsbeschränkte Blockchain haben in diesem Szenario Nutzer keinen Zugriff auf den Datenbestand und somit auch keine Möglichkeit, die Integrität selbst zu prüfen. Durch das Veröffentlichen des Hashwerts und dem Hinterlegen eines smart contracts wäre eine eingeschränkte Validierungsmöglichkeit, auch unabhängig eines Eingreifens der Archiveinrichtungen, gegeben. Auch im digitalen Magazin muss es die Möglichkeit geben Objekte zu löschen, sei es aus regulatorischen Gründen, zum Schutz der Rechte Dritter oder bei menschlichen Fehlern. Prinzipiell ist die Löschung in einer Blockchain zwar nicht vorgesehen, dennoch ist über Umwege eine entsprechende Architektur möglich. Erst jüngst stellte der bekannte Records Manager Dr. Kampffmeyer sein Konzept für das Löschen oder Ersetzen von Informationen in einer Blockchain vor. Dieser Ansatz sieht vor, sämtliche Metadaten und Hashwerte eines Informationsobjekts über ein Audit-Trail – ein System zur Prüfung von Prozessen – zu führen und dieses Audit-Trail in eine Blockchain zu implementieren. Über eine Fernsignatur, so Kampffmeyer, wäre auch eine zusätzliche qualifizierte Signierung nach eIDAS möglich. Bei einer Löschung würden die Metadaten des Objekts sowie des Vorgänger- und Nachfolgerobjekts ausgelesen, der Grund für die Löschung angegeben und schließlich das Objekt entweder logisch oder physisch gelöscht.²⁶ Diese ineinander integrierte Lösung erhöht wesentlich den Komplexitätsgrad, bietet jedoch dadurch die Chance, die unterschiedlichen Intentionen der DSGVO, eIDAS und OAIS miteinander zu harmonisieren.²⁷

25 Vgl. Collomosse, John et. al.: ARCHANGEL, S. 3f.

26 Vgl. Kampffmeyer: Konsistentes Löschen in der Block Chain. Hamburg 2019, S. 2ff. URL: https://www.project-consult.de/files/20190425_Konsistentes_Loeschen_in_der_Block_Chain_Kff.pdf [Abruf 4. 10. 2019].

27 Vgl. für Lösungsansätze auch: Schwalm, Steffen: Neue Besen im Spannungsfeld eIDAS und DSGVO-Blockchain für (dauerhafte) Verzeichnungsdienste? Berlin 2019, F. 19ff. URL: https://www.researchgate.net/publication/330856590_Neue_Besen_im_Spannungsfeld_eIDAS_und_DSGVO-Blockchain_fur_dauerhafte_Verzeichnisdienste; Lemieux, Victoria: Trusting Re-

Risiken und Chancen

Neue Technologien und Paradigmen bringen stets neue Probleme und Herausforderungen mit sich. So ist eine Blockchain als native Einzellösung schwierig zu implementieren, will man sich nicht auf Standardkomponenten verlassen. Gleichzeitig gibt es noch ungeklärte Fragen, deren Beantwortung derzeit erst erarbeitet wird: Wie ist die rechtliche Anerkennung? Wie lässt sich das Ledger mit den derzeitigen IT-Standards vereinbaren? Lässt sich die Blockchain standardisieren, um sie professionell in Unternehmen und der öffentlichen Verwaltung produktiv einzusetzen?²⁸ Als ein aktuelles Hype-Thema kann Blockchain in einigen Jahren auch wieder obsolet sein. Die Themen Datenschutz und der IT-Sicherheit werden in den Folgejahren einen immer breiteren Raum einnehmen. Dementsprechend ist es kritisch zu sehen, inwieweit eine echte Blockchain datenschutzkonform eingerichtet werden kann, ohne ihre inhärente Sicherheit durch hohe Eintrittshürden und eine Vielzahl an Teilnehmern einzubüßen. Mit dem Einsatz einer Blockchain sind aber nicht nur Risiken verbunden, sondern auch viele Chancen. Die authentische und integre Erhaltung digitaler Informationen kann durch Blockchain ein neues, professionelles Level erreichen. Ist bei einer Prüfung eine Transaktion als korrumpiert aufgefallen, so sind alle folgenden Elemente, bzw. sogar das ganze Netzwerk als dementsprechend korrumpiert anzusehen. Anders hingegen kann durch eine positive Prüfung die Integrität des gesamten Netzwerks nachgewiesen werden. Insbesondere das bisher im Archivwesen praktizierte „Ius Archivi“ – die inhärente und institutionalisierte Vertrauenswürdigkeit eines Archivs bzw. des übernommenen Archivgutes²⁹

.....
 cords. Is Blockchain technology the answer? In: Records Management Journal. 26(2016), S. 110–139. URL: <https://doi.org/10.1108/RMJ-12-2015-0042>; Keller-Marxer, Peter: Blockchains als Herausforderung und Instrument des Records Management. Kritische Würdigung verteilter, kryptographisch gesicherter Transaktionsjournale in Bezug auf die Vertrauenswürdigkeit digitaler Records. Masterarbeit Universität Bern. Bern 2016. URL: https://www.researchgate.net/profile/Peter_Keller-Marxer/publication/313875070_Blockchains_als_Herausforderung_und_Instrument_des_Records_Management_-_Kritische_Wurdigung_verteliter_kryptographisch_gesicherter_Transaktionsjournale_in_Bezug_auf_die_Vertrauenswurdigkeit_digitaler_/links/58ac5d33aca272c47d9d342f/Blockchains-als-Herausforderung-und-Instrument-des-Records-Management-Kritische-Wuerdigung-verteilter-kryptographisch-gesicherter-Transaktionsjournale-in-Bezug-auf-die-Vertrauenswuerdigkeit-digital.pdf [Abruf 4. 10. 2019].

28 Für aktuelle Arbeiten zur Standardisierung & Normung siehe u. a. den Technical Report der DIN-SPEC 3104:2019-04: Blockchainbasierte Datenvalidierung. Berlin 2019. URL: https://www.researchgate.net/profile/Steffen_Schwalm/publication/332029196_DIN_SPEC_3104_Blockchainbased_Data_ValidationBlockchainbasierte_Datenvalidierung_/links/5caf6ed54585156cd7914e6b/DIN-SPEC-3104-Blockchainbased-Data-Validation-Blockchainbasierte-Datenvalidierung.pdf?origin=publication_detail;

EU-Blockchain Observatory and Forum: Blockchain and digital Identity. Brüssel 2019. URL: https://www.euBlockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf?width=1024&height=800&iframe=true [Abruf 4. 10. 2019].

29 Vgl. Schäfer, Udo, Authentizität. Elektronische Signaturen oder Ius Archivi? In: Schäfer, Udo; Hering, Rainer (Hrsg.): Digitales Verwalten – Digitales Archivieren. Hamburg 2004 (=Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg, Bd. 19) S. 13-31, hier S. 26ff.; Schäfer, Udo: Authentizität. Vom Siegel zur elektronischen Signatur. In: Archivierung elektronischer Unterlagen. Stuttgart 1999 (=Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Bd. 13), S. 165-181, hier S. 169ff.

erscheint mit dieser Methodik vergleichbar. Die Umsetzung eines per se vertrauenswürdigen digitalen Archivs nur mittels der Blockchain-Technologie greift jedoch zu kurz. Ein Ledger kann sicher auf technischer Ebene in Teilbereichen hierzu beitragen; ein ganzheitliches Konzept, basierend auf der DIN 31644 bzw. der entsprechenden ISO 16363, ersetzt es nicht. Eine Ledger ist es im technischen Rahmen sehr flexibel und anpassbar. Man darf dabei allerdings die eigentlichen Grundprinzipien nicht außer Acht lassen. So stellt eine Blockchain-Lösung, die lediglich einen einzelnen Teilnehmer aufweist, bereits ein gewisses Persiflieren dar und ist nicht im Sinne des unabhängigen Integritätsnachweises von außerhalb.

Viele Akteure beschäftigen sich mit dem Thema Blockchain. Hierbei soll nicht unerwähnt bleiben, dass vieles davon natürlich auf den genannten Hype zurückgeht. Die deutsche Bundesregierung plant bspw. eine nationale Blockchain-Strategie bis Mitte 2019. Das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) legte hierzu ein erstes Grundgutachten vor.³⁰ Neben der Evaluierung von Ergebnissen einzelner Förderprojekte sollen grundlegende Erkenntnisse von Behörden und Experten über eine Online-Konsultation miteinfließen. Ein wichtiger Themenkomplex, dem sich die Bundesregierung widmen möchte, ist die Begutachtung der Blockchain im Kontext des europäischen und nationalen Datenschutzes und der Beantwortung von rechtlichen Fragen im Umgang mit der Blockchain. Das deutsche Bundesamt für Sicherheit in der Informationstechnik (BSI) präsentierte im Mai 2019 auf der 16. IT-Sicherheitskonferenz einen ergänzenden Blockchain-Leitfaden.³¹

Fazit

Der Beweiswert und die vertrauenswürdige Bewahrung von Informationen werden uns gesellschaftlich wie auch fachspezifisch noch die nächsten Jahre und Jahrzehnte beschäftigen. Blockchain basiert auf Ansätzen, die in der IT grundlegend bereits seit längerem bekannt sind, und integriert diese in eine innovative Infrastruktur. Der wesentliche Vorteil liegt in der jeweils entstehenden Community, die statt Behörden und Vertrauensdiensten eigenständige Lösungen anbieten kann, um aus sich heraus einen Beweiswert und die Vertrauenswürdigkeit nachzuweisen. Für viele Bereiche, wie der allgemeinen Kommunikation im Internet, ist die Blockchain nicht die passende Lösung oder die derzeitigen Lösungen werden als ausreichend erachtet. Für diejenigen Bereiche mit erhöhtem Bedarf an Vertrauenswürdigkeit, wie die digitale Archivierung, ist die Blockchain-Technologie jedoch ein interessanter Ansatz. Wie die meisten Hype-Themen ist die Blockchain dabei kein Heilsbringer, es bringt für den Integritätsnachweis und der Vertrauenswürdigkeit des nutzenden Verbundes

30 Bundesministerium für Verkehr und digitale Infrastruktur: BMVI präsentiert Blockchain-Gutachten. URL: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/Blockchain-grundgutachten.html?nn=12830> [Abruf 4. 10. 2019].

31 Bundesamt für Sicherheit in der Informationstechnik: Blockchain sicher gestalten. Konzepte, Anforderungen, Bewertungen. Bonn 2019. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Krypto/Blockchain_Analyse.pdf?__blob=publicationFile&v=5#download=1 [Abruf 4. 10. 2019].

neue Lösungen. Andere Themen werden hingegen nicht angegangen oder neue Herausforderungen geschaffen. Die noch fehlende Standardisierung und der datenschutzkonforme Umgang stellen die größten Herausforderungen beim Einsatz eines Ledgers in der digitalen Archivierung dar. Durch die aktuelle Forschung werden allerdings praktikable Lösungsansätze vorgestellt und erarbeitet. Unabhängig von einer spezifischen Technologie müssen sich die Mitglieder der Fachcommunity dabei fragen, was brauchen wir eigentlich? Die fachlichen Anforderungen an den Beweiswert und die Vertrauenswürdigkeit müssen für die zukünftige Bewahrung definiert und am jeweiligen Stand der Technik immer wieder neu kritisch evaluiert werden. Der Integritätsnachweis im Verbund, der rechtssichernde Nachweischarakter der digitalen Archivierung u. v. m. sind Themen, die in Anbetracht zukünftiger Ansprüche seitens des eGovernments, der Nutzer und nicht zuletzt auch der Community der digitalen Archivierung selbst neu und allumfänglich aufgearbeitet werden müssen.³²

Quellen

- Antonopoulos, Andreas: Bitcoin & Blockchain. Grundlagen und Programmierung. Die Blockchain verstehen, Anwendungen entwickeln. 2. Aufl. Heidelberg 2018.
- Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V.: Online-Konsultation zur Erarbeitung der Blockchain-Strategie der Bundesregierung. Bitkom-Stellungnahme. Berlin 2019. URL: https://www.bitkom.org/sites/default/files/2019-04/190408_stellungnahme_Blockchain-strategie-konsultation_online.pdf [4. 10. 2019].
- Bundesamt für Sicherheit in der Informationstechnik: Blockchain sicher gestalten. Konzepte, Anforderungen, Bewertungen. Bonn 2019. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Krypto/Blockchain_Analyse.pdf?__blob=publicationFile&v=5#download=1 [Abruf 4. 10. 2019].
- Bundesministerium für Verkehr und digitale Infrastruktur: BMVI präsentiert Blockchain-Gutachten. URL: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/Blockchain-grundgutachten.html?nn=12830> [Abruf 4. 10. 2019].
- Bundesverband IT-Sicherheit: TeleTrust-Positionspapier "Blockchain". Handreichung zum Umgang mit der Blockchain. Berlin 2017. URL: https://www.teletrust.de/fileadmin/docs/publikationen/broschueren/Blockchain/2017_TeleTrust-Positionspapier_Blockchain__.pdf [Abruf 4. 10. 2019].
- Burgwinkel, Daniel (Hg.): Blockchain Technology. Einführung für Business und IT-Manager. Berlin, Boston 2016.
- Collomosse, John et. al.: ARCHANGEL: Trusted Archives of Digital Public Documents. In: ACM Document Engineering. o. O. 2018. URL: <https://arxiv.org/abs/1804.08342> [Abruf 4. 10. 2019].
- DIN-SPEC 3104:2019-04: Blockchainbasierte Datenvalidierung. Berlin 2019. URL: https://www.researchgate.net/profile/Steffen_Schwalm/publication/332029196_DIN_SPEC_3104_Blockchainbased_Data_Validation-Blockchainbasierte_Datenvalidierung/links/5caf6ed54585156cd7914e6b/

.....
 32 Zur Evaluation des Beweiswertes vgl. auch Keller-Marxer, Peter: Blockchains als Herausforderung für das Records Management, S. 296.

- DIN-SPEC-3104-Blockchainbased-Data-Validation-Blockchainbasierte-Datenvalidierung.pdf?origin=publication_detail [Abruf 4. 10. 2019].
- Ebersbach, Jan: Blockchain-Anwendungen. Blockchain im praktischen Einsatz. In: iX. Magazin für professionelle Informationstechnik, Nr. 7 (2018), S. 50-54.
 - EU-Blockchain Observatory and Forum: Blockchain and digital Identity. Brüssel 2019. URL: https://www.euBlockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf?width=1024&height=800&iframe=true [Abruf 4. 10. 2019].
 - Glatz, Florian; Holthusen, Jannis; Kufeld, Simon: Vorstellung der Blockchain-Technologie. o. O. 2016. URL: <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Vorstellung%20der%20Blockchain-Technologie.pdf> [Abruf 4. 10. 2019].
 - Kampffmeyer: Konsistentes Löschen in der Block Chain. Hamburg 2019, S. 3. URL: https://www.project-consult.de/files/20190425_Konsistentes_Loeschen_in_der_Block_Chain_Kff.pdf [Abruf 4. 10. 2019].
 - Keller-Marxer, Peter: Blockchains als Herausforderung und Instrument des Records Management. Kritische Würdigung verteilter, kryptographisch gesicherter Transaktionsjournale in Bezug auf die Vertrauenswürdigkeit digitaler Records. Masterarbeit Universität Bern. Bern 2016. URL: https://www.researchgate.net/profile/Peter_Keller-Marxer/publication/313875070_Blockchains_als_Herausforderung_und_Instrument_des_Records_Management_-_Kritische_Wurdigung_verteilter_kryptographisch_gesicherter_Transaktionsjournale_in_Bezug_auf_die_Vertrauenswurdigkeit_digitaler_/links/58ac5d33aca272c47d9d342f/Blockchains-als-Herausforderung-und-Instrument-des-Records-Management-Kritische-Wuerdigung-verteilter-kryptographisch-gesicherter-Transaktionsjournale-in-Bezug-auf-die-Vertrauenswuerdigkeit-digital.pdf [Abruf 4. 10. 2019].
 - Keller-Marxer, Peter: Blockchains als Herausforderung für das Records Management. In: Informationswissenschaft: Theorie, Methode und Praxis 5 (2018), S.268-297. URL: <https://doi.org/10.18755/iw.2018.30> [Abruf 4. 10. 2019].
 - Lemieux, Victoria: Trusting Records. Is Blockchain technology the answer? In: Records Management Journal. 26 (2016), S. 110-139. URL: <https://doi.org/10.1108/RMJ-12-2015-0042> [Abruf 4. 10. 2019].
 - Nakamoto, Satoshi: Bitcoin. A Peer-to-Peer Electronic Cash System. o. O. 2008. URL: <https://bitcoin.org/bitcoin.pdf> [Abruf 4. 10. 2019].
 - Roth, Matthias: Wie funktioniert eine Blockchain? o. O. 2017. URL: <https://www.adesso.de/de/news/blog/wie-funktioniert-eine-Blockchain.jsp> [Abruf 4. 10. 2019].
 - Schäfer, Udo: Authentizität. Vom Siegel zur elektronischen Signatur. In: Archivierung elektronischer Unterlagen. Stuttgart 1999 (=Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Bd. 13), S. 165-181.
 - Schäfer, Udo, Authentizität. Elektronische Signaturen oder Ius Archivi? In: Schäfer, Udo; Hering, Rainer (Hrsg.): Digitales Verwalten – Digitales Archivieren. Hamburg 2004 (=Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg, Bd. 19) S. 13-31.
 - Schwalm, Steffen: Neue Besen im Spannungsfeld eIDAS und DSGVO-Blockchain für (dauerhafte) Verzeichnungsdiene? Berlin 2019. URL: <https://www.schwalm.com/de/wordpress/wp-content/uploads/2019/07/Blockchain-fuer-Verzeichnungsdiene.pdf> [Abruf 4. 10. 2019].

[//www.researchgate.net/publication/330856590_Neue_Besen_im_Spannungsfeld_eIDAS_und_DSGVO-Blockchain_fur_dauerhafte_Verzeichnisdienste](https://www.researchgate.net/publication/330856590_Neue_Besen_im_Spannungsfeld_eIDAS_und_DSGVO-Blockchain_fur_dauerhafte_Verzeichnisdienste) [Abruf 4. 10. 2019].

- Seeger, Jürgen: Von Bitcoin bis Ripple. Hyperledger-Implementierungen im Vergleich. In: iX. Magazin für professionelle Informationstechnik, Nr. 7 (2018), S. 44-48.
- Tapscott, Alex; Tapscott, Don: Die Blockchain-Revolution. Wie die Technologie hinter Bitcoin nicht nur das Finanzsystem, sondern die ganze Welt verändert, Wiesbaden 2016.
- Verordnung (EU) Nr. 910/2014 des Europäischen Parlaments und des Rates über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG [eIDAS-VO]. Brüssel 2014. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32014R0910&from=DE> [Abruf 4. 10. 2019].
- Wieler, Herbert: Blockchain. Was ist der Unterschied zu einer herkömmlichen Datenbank? o. O. 2018. URL: <https://www.infopoint-security.de/Blockchain-was-ist-der-unterschied-zu-einer-herkoemmlichen-datenbank/a15631/> [Abruf 4. 10. 2019].

Summary

Blockchain is currently experiencing a media hype. It claims to have the potential to sustainably shape Internet-based communication through the use of cryptographic techniques and even revolutionize the Internet. In particular, the high expectations in a high degree of authenticity and integrity of the transactions carried out and the managed data in a Blockchain make this technology appear interesting from the point of view of long-term preservation. The Blockchain – or Distributed-Ledger-Technology (DLT) – represents a digital protocol register in a Peer-to-Peer-network in which users exchange data and also validate each others data via a consensus mechanism. The project Archangel of the National Archives UK represents the first major productive implementation of a Blockchain in an OAIS-compliant digital archive as well as in a GLAM institution. A comparison of the Blockchain functions in the established national and international standards and norms such as DIN 31644 and eIDAS shows that a classic Blockchain is not compatible with the previous standards of digital archiving. Likewise the conformity to the DSGVO is not given with a classical Blockchain. By individually adapted solutions, however, conformity can be achieved. The current work on standardization as well as the research show, among other things, on the basis of a combined audit trail solution in a Blockchain, a way that would be both compliant to OAIS and to eIDAS. In conclusion, it can be said that the requirements for evidence and trustworthiness must always be critically evaluated using the state of the art in the latest technology.

VOM UMGANG MIT FEHLERHAFTEN DATEN IM INGEST

Ein Werkstattbericht aus der Entwicklung der DA-NRW Software Suite (DNS).

Jens Peters

Vom Ausnahmefall zum „Normalfall“ im Ingest?

Analog zur Arbeitsweise eines klassischen Archivs kann ein digitales Archiv archivwürdige aber nur beschränkt archivfähige Dateien nicht zurückweisen. Ein analoges Schriftstück hoher archivarischer Relevanz mit Beschädigung wird von einem Archiv selbstverständlich übernommen und fachgerecht konservatorisch behandelt und entsprechend gesichert.

In der Praxis der digitalen Archivierung am Beispiel der „DA NRW Software Suite (DNS)“ zeigt sich hingegen ein Dilemma auf, für das die Domäne insgesamt Lösungen finden muss, die auch bei der zu erwartenden hohen Datenmenge praktikabel einsetzbar sind: Lesefehler und Zugriffsfehler bei Dateioperationen sind praktisch nicht vorgesehen. Die digitalen Systeme gehen von „perfekten Dateien“ aus. Die Wirklichkeit sieht hingegen anders aus.

Oftmals ist aber gerade die Bewertung und die Bemessung des Schweregrads der Beschädigung eine komplexe Fragestellung, bei der es auf den zweiten Blick keine gute oder gar richtige Antwort geben mag. Sie muss nicht notwendigerweise in einen objektivierbaren Sinn „korrekt“ sein. Zum einen verläuft die Entwicklung von Werkzeugen, die diese Erkennung mitunter leisten, sprunghaft. Zum anderen ist die Aussage des benutzten Werkzeuges selbst eine Interpretation von Standards oder Normen zu einem bestimmten Zeitpunkt. Die Interpretation muss also mitnichten korrekt sein.

Im Konfliktfeld aus Zurückweisung oder der massenhaften Übernahme von ungeprüftem Material im Sinne einer einfachen *bitstream*-Archivierung wird im Folgenden kurz aus der Praxis der Arbeit mit der „DA NRW Software Suite“ (DNS) berichtet. Die Erfahrungen resultieren aus Probeübernahmen von Produktionsdaten zukünftiger Servicenehmer, die z. T. eine Datenmenge von jeweils bis zu 60 GB pro Tag avisieren.

Es wird hier ein Lösungsszenario skizziert, welches die einzelnen Schritte der Ingestverarbeitung bestimmten Qualitätskategorien zuordnet, die aufeinander aufbauen. Jede Kategorie wird dabei von einem Qualitätsmerkmal gekennzeichnet, welche Ausdruck der Eignung als digitales Archivgut ist. Diese ermöglichen eine angemessene Aufbewahrung und Verwaltung auch problematischer Dateien, bis beispielsweise ein Datenarchivar/Datenarchäologe bzw.

Preservation Manager sich der beschädigten Daten genauer und zielgerichtet annehmen kann. Alle Zuordnungen zu Qualitätskategorien werden für den Servicenehmer transparent vollzogen und bekommen einen nachvollziehbaren Eintrag in den Langzeitarchivierungsmetadaten.

Diese Qualitätskategorien sind prototypisch für die Langzeitarchivierungslösung „DA NRW Software Suite“ implementiert und befinden sich gerade in der Übernahmephase in den Echtbetrieb.

Die „DA NRW Software Suite“ ist ein OAIS konformes Langzeitarchivierungssystem mit mehrfach redundanten, dezentralen Archivspeichern und zentralem Erhaltungsmanagement (preservation planning). Als besondere Erweiterung wird für entsprechende veröffentlichungsfähige Objekte eine Darstellung in einer Präsentationsschicht ermöglicht, die über Schnittstellen diese Objekte zukünftig auch in nationale und internationale Portale weitergeben kann.

Die DA NRW Software Suite stellt das *quelloffene* Lösungsmodul des Lösungsverbandes¹ des Digitalen Archivs in Nordrhein-Westfalen (DA NRW) dar. Der gemeinsame Verantwortungsbereich von Land und Kommunen wird von einer Arbeitsgemeinschaft DA NRW, bestehend aus dem Land NRW, vertreten durch das Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen und dem KDN² getragen, der ein Beirat in beratender Funktion und eine Geschäftsstelle in unterstützender Funktion angehören³. LVR-InfoKom, als Systemhaus des Landschaftsverbandes Rheinland (LVR)⁴ in Köln, entwickelt die Softwarelösung DNS im Auftrag weiter.

Die DNS verfügt über eine redundante Datenhaltung an drei geografisch voneinander getrennten Speicherlokationen und ein zentrales Erhaltungsmanagement. Neben der Entwicklungspatenschaft ist LVR-InfoKom ein Betreiber eines Speicherknotens im Verbund mit anderen Partnern im Land. Die Software versteht sich als Langzeitarchivierung „as a service“ und steht grundsätzlich allen Kultur, Gedächtnis- und Wissenschaftseinrichtungen in Nordrhein-Westfalen zur Nutzung bereit.

„Technisch problematisch“ – Eine Einordnung

Die Einordnung ob eine einzuliefernde Datei „technisch problematisch“ ist, ist eine Fragestellung bei der es bei näherer Betrachtung keine eindeutig korrekte Antwort geben mag, wie einleitend schon formuliert wurde. Die Bewertung ob und wann eine Datei defekt ist, ist von vielen Faktoren abhängig. Technisch betrachtet handelt es sich zunächst um einen Lesefehler während der Interpretation der gespeicherten Daten, also um einen Abbruch oder eine Warnung beim Zugriff auf die Datei. Diese Feststellung ist wichtig, denn es handelt sich **nicht** um Korrumpierung der Daten auf dem Transportweg. Hierzu gibt es bereits sehr gute Lösungsvorschläge⁵.

.....
1 Neben dem weiteren Lösungsmodul DiPS.kommunal.

2 Dachverband kommunaler IT-Dienstleister in NRW

3 Mehr zu DA NRW unter <http://www.danrw.de> [Abruf 29.04.19]

4 Mehr zum Landschaftsverband Rheinland unter <http://www.lvr.de> [Abruf 29.04.19]

5 Vgl. z. B. den Standard „BagIt“ der Library of Congress unter <https://sourceforge.net/>

Als konkretes Problem aus der Praxis sei hier eine TIFF Datei mit einem vom Bildverarbeitungsprogramm des DNS Systems als fehlerhaft vermerkten Eintrag in einem IPTC Erweiterungsfeld genannt. Zwischen zwei Versionsnummern des Bildbearbeitungswerkzeuges ImageMagick hat sich die Interpretation von „Datei ist fehlerhaft“ zu „Datei ist lesbar, produziert aber eine Warnung“ geändert. Ein weiteres Beispiel könnte eine PDF-A Datei sein, die bei der Validitätsprüfung mit verschiedenen Validatoren ein unterschiedliches, gar widersprüchliches Ergebnis produziert.

Oftmals sind auch die installierten Komponenten und deren Konfiguration mitverantwortlich für die Einordnung als „problematisch“. Dadurch, dass sich die benutzten Werkzeuge und auch die verwendeten Betriebssystemkomponenten ständig weiterentwickeln, rückt hier auch der Zeitpunkt der Prüfung in den Fokus. Ist eine Datei, die heute nicht korrekt ausgewertet werden kann, morgen noch ebenso unleserlich? Ist ein angezeigter Lesefehler der Datei in einer der Routine zur Ermittlung des Dateiformats (z. B. Fido) nun ein Fehler der Datei oder der möglicherweise veralteten Version des Programms? Handelt es sich vielleicht um einen strukturellen Unterschied in den verwendeten Systemen auf Sender oder Empfängerseite? Kann die Datei jetzt oder in Zukunft repariert werden? Dies sind nur einige Fragen, die sich in diesem Zusammenhang stellen ließen. Zur Strukturierung dieser Fragestellungen sollen hier im Folgenden unterschiedliche Perspektiven beleuchtet werden.

Aus Sicht des OAIS Referenzmodells

Aus Sicht des OAIS Referenzmodells ist zunächst ein „*review committee*“⁶ für problematische Dateieinlieferungen zuständig, welches den Fehler prüfen und bewerten soll. Es kann entscheiden, ob es die Datei in Folge trotzdem übernimmt, ganz ausschließt oder eine Neueinlieferung verlangt. Eine Neueinlieferung ist aber nur unter der Prämisse sinnvoll, wenn diese das Problem auch tatsächlich löst. Dies wird nur in dem Fall möglich sein, wenn der Einliefernde durch die Anbietetung selbst das Problem erst verursacht hat – das Abgabeobjekt (das SIP oder eine einzelne Datei) also noch in einem korrekten Zustand beim Einliefernden greifbar oder überhaupt vertretbar korrigierbar ist. Oft ist dies aber nicht der Fall. Diese Problemstellung soll im Folgenden kurz skizziert werden.

Aus Sicht der abgebenden Stelle

Es erscheint logisch, dass ein digitales Archiv eine Neueinlieferung eines Abgabepakets (SIP) verlangt, wenn dies durch eine Korrumpierung des Abgabepakets selbst (SIP) entstanden ist. Ursachen könnten hier eine fehlerhafte Erzeugung oder eine Korrumpierung auf dem Transportweg sein. Allgemeiner gesagt, hierunter sind Verstöße gegen die Integrität und Authentizität zu verstehen. Für diese Fälle sind bereits Lösungen vorgeschlagen worden. Diese lassen sich relativ einfach in einen automatischen Prüfungsprozess einbauen⁷. Ebenso können SIP, die nicht der abgesprochenen Bildungsregel von SIP gehorchen

.....
 projects/loc-xferutils/ [Abruf 29.04.19]

6 Vgl. S.13 “Reference Model for an Open Archival Information System (OAIS)”, The Consultative Committee for Space Data Systems (CCSDS), 2012, unter: <https://public.ccsds.org/pubs/650x0m2.pdf> [Abruf 19.02.18]

7 Vgl. z. B. den Transferstandard BagIt oder XDOMEA 2.2

durch einfache Neuerzeugung gelöst werden. Schwierig bleibt der übrige Teil der problematischen Dateien:

Die abgebende Stelle kann eventuell die Datei gar nicht mehr in einer korrekten Variante liefern, da die Werkzeuge zur Erzeugung einer derartigen Datei gar nicht mehr vorliegen oder das Wissen über die Erstellung der Datei nicht mehr vorhanden ist. Meist kann die abgebende Einrichtung den Fehler nicht oder nicht mit vertretbarem Aufwand korrigieren oder darf dies eventuell auch gar nicht⁸. Hierbei würde eine einfache Neuversendung aus Sicht der abgebenden Stelle **zum identischen Resultat beim digitalen Archiv** führen. Wenn es sich um automatische Übernahmen handelt, könnte ggfs. eine hohe Datenmenge zur Abnahme anstehen, Verzögerungen in der Abnahme der Datenmenge führen unweigerlich zu Unmut bei der abgebenden Institution.

Aus Sicht des digitalen Archivs

Wie schon gezeigt wurde, ist die Frage, ob die von einem digitalen Archiv bemängelten Dateien überhaupt in einem objektiven Sinne „kaputt“ sind, nicht so einfach zu beantworten. Verweigert hier das digitale Archiv die Verarbeitung, führt das zu berechtigtem Unmut bei der abliefernden Stelle, die ihre Dateien nicht entsprechend abgeben kann. In der Praxis entzündeten sich Diskussionen zumeist an Fragen nach der Bewertung des Fehlers selbst – vor allem wenn die Datei grundsätzlich lesbar erscheint. Insbesondere bekannte Schwachstellen auch etablierter Werkzeuge (z. B. JHOVE⁹) oder der interpretationsfreudigen Implementation von Formatspezifikationen (z. B. TIFF) in verschiedenen Programmbibliotheken können zusätzlich zu Verwirrungen führen¹⁰. Zudem sind Archiv und abgebende Stelle auf eine zügige Abarbeitung insbesondere bei Massendaten auch aus Zeit- und Speicherplatzgründen angewiesen.

Einige Systeme haben hierfür „Überarbeitungspostkörbe“¹¹ oder Benachrichtigungssysteme bei Fehlern bereits implementiert¹². Diese erlauben zumeist die Speicherung als „*bitstream*“ inklusive aller bisher angefallener Verarbeitungsdaten und der Bereitstellung eines Arbeitsberichts, nach Sichtung durch einen Mitarbeiter des Archivs. Dies erscheint aus zweierlei Dingen nicht praktikabel. Zum einen verschiebt es die Verarbeitung in die Zukunft der weiteren digitalen Bestandserhaltung, ohne diese Objekte hinsichtlich ihrer potentiellen Eignung nach dem Stand der Technik eindeutig zu qualifizieren. Zum anderen erscheint dies bei Massendaten aus Retrodigitalisierungsprozessen und aus der Menge der zu erwartenden „*born digitals*“, die in der nahen Zukunft zur Archivierung anfallen werden, als kaum machbar.

Bestandserhaltende Prozesse kann das digitale Archiv ebenfalls nicht zielgerichtet auf die problematischen Dateien hin entwickeln – es hat haufenweise

8 Bspw. bei der perspektivischen Archivierung von Dokumenten, die dem Pflichtabgabegesetz unterlägen.

9 Zu JHOVE <http://jhove.openpreservation.org/> [Abruf 02.05.19]

10 z. B. bei Nutzung von IPTC-Tag-Thematik bei TIFF durch unterschiedliche/fehlerhafte Implementationen der vorgeschlagenen Standards.

11 z. B. das DA NRW Lösungsmodul „DIPS.kommunal“

12 z. B. das DA NRW Lösungsmodul „DNS“, vgl. Abschnitt 3.

ungeprüfte Daten übernommen, für die es allenfalls *bitstream preservation* zu-sagen kann.

Im Detail erweisen sich diese sehr technischen Fragen als Fragen von erheblicher Relevanz und Ausmaß, die erhebliche Aufwände produzieren können. Für das erstmalige Auftreten eines solchen Fehlers macht die Analyse durch ein „*review commitee*“ sicherlich sinnvoll erscheinen – in der Folge muss es aber Prozeduren für eine möglichst einfache Übernahme geben. Der spezielle Fehler muss also für das Digitale Archiv einfach verwaltbar sein und für spätere daten-archäologische Behandlung.

Eine Lösung aus diesem Dilemma könnte die Verwendung von durchgängig zu verwendenden „**Qualitätskategorien**“ in der digitalen Archivierung sein.

Da die „Schwere“ eines Fehlers oft nicht messbar ist und auch durch unterschiedliche Auslegung von Formatspezifikationen ein gewisser Interpretations-spielraum für die Bezeichnung als ein „Fehler“ insgesamt gegeben ist, wird eine Einstufung nach einer abstrakten Skala von „Eignung zur Digitalen Langzeitarchivierung („Qualitätskategorien“)“ vorgeschlagen, mittels derer das Qualitätsmanagement technisch unterstützt werden kann.

Qualitätskategorien bei der DNS

Es erscheint sinnvoll, Qualifikationsmerkmale zum Ausdruck der Eignung als digitales Archivgut im Übernahme/Ingestprozess eines Systems zur digitalen Langzeitarchivierung zu implementieren, also in der Phase, in der die eintreffenden SIP in AIP umgeformt werden. Durch die transparente Festlegung und der verbindlichen Spezifikation der Kriterien für jede Kategorie werden zielgerichte-te, bestandserhaltende Maßnahmen für das digitale Archiv ebenso möglich, wie eine Transparenz der Entscheidung für den Servicenehmer. Auch bei späteren Ergänzungen von Archivpaketen durch Einsendung eines so genannten „Deltas“ ist ein Kategoriewechsel, sowohl in die eine (eine „Herabkategorisierung“), als auch in die andere Richtung (zu einer „höheren“ Kategorie) durchaus möglich. Gleiches trifft auf bestandserhaltende Maßnahmen zu. Die folgenden Kategorien sind vorgesehen:

Kategorie 0

Abgabepakete der Kategorie 0 sind schlicht keine validen SIP im Sinne des OAIS und der Vorgaben des digitalen Archivs, daher ist die Kategorie 0 keine Stufe zur technischen Qualifizierung von **Archivobjekten** des digitalen Archivs im eigentlichen Sinn. SIP der Kategorie 0 können unter keinen Umständen erfolgreich zu AIP umgeformt werden. Die Abgabepakete gehorchen nicht der SIP Spezifikation der DNS. Beim Transport korrumpierte SIP sind nicht annehmbar und werden auch nicht vom DNS aufgenommen. Eine Neuversendung durch den Servicenehmer ist in jedem Fall sinnvoll und wird in der Regel das Problem auch lösen.

Kategorie 1

Pakete der Kategorie 1 haben mindestens die Kriterien SIP Konformität und (ggfs.) Virenfreiheit zu erfüllen. Sie sind auf dem Transportweg zwischen dem Ablieferer und der Übergabe an das elektronische Archiv nicht korrumpiert

worden, die Authentizität und Integrität kann zweifelsfrei festgestellt werden. Des Weiteren kann aktuell keine bekannte Schadsoftware in dem SIP identifiziert werden. Ab dieser Stufe kann zumindest eine „*bitstream preservation*“ des reinen Datenstroms gewährleistet werden. Bestandserhaltende Prozesse für Pakete dieser Art können aber Stand heute vom Digitalen Archiv ebenso wenig zugesagt werden, wie die Erstellung von Derivaten für andere Anzeigekontexte (z. B. die automatische Erzeugung von Ansichtsvarianten der Dateien).

Kategorie 2

Fehlerfreie Formatidentifikation anhand nach dem Stand der Technik, z. B. mittels FIDO, DROID und PRONOM Identifizieren und ggfs. Codeerkennung mittels weiterer Programmbibliotheken, gelten als erfolgreich gelungen. Bestandserhaltende Maßnahmen sind ab dieser Kategorie möglich.

Kategorie 2a)

Die Dateien sind nach Aussage der benutzten Formatvalidierer valide. Insbesondere die Kombination aus Spezialvalidieren je erkanntem Dateityp scheint auf Grund der anschwellenden und korrekten Kritik an JHOVE allenthalben geboten¹³. Aber eine Konvertierung zu langzeitstabilen Formaten ist nach dem heutigen Stand der Technik nicht erfolgreich verlaufen. Nach heutigem Stand der Technik können bestandserhaltende Maßnahmen nur für diejenigen Dateien zugesagt werden, für die eine Konvertierung erfolgreich verlaufen ist.

Kategorie 2b)

Eine Datei ist auf Grund der verwendeten Standard- und Spezialformatvalidierer nicht valide. Aber die Wandlung in langzeitstabile Format ist erfolgreich. Als Resultat liegt ein fehlerfreies und validierbares Konvertierungsprodukt in einem langzeitstabilen Format vor. Für die Kategorie 2b können bestandserhaltende Prozesse nach derzeitigem Stand der Technik zugesagt werden.

Kategorie 3 und 4 (und ggfs. weitere)

Alle vorhergehenden Kategorien sind erfolgreich erfüllt worden, zusätzlich sind systematische, eigene Prüfung auf Grund weiterer Kriterien zur Messung von Qualität erfüllt worden. Beispielsweise wurden aus Sicht des digitalen Archivs ausschließlich langzeitstabile Formate eingeliefert, oder eine notwendige Eingangskonvertierung in entsprechend langzeitstabile Formate war für alle Formate erfolgreich. Nach dem Stand der Technik ist aber eine Langzeiterhaltung möglich.

Die Kategorien bauen weitestgehend aufeinander auf, die jeweils Folgenden enthalten die Vorgegangenen. Wichtig ist auch der Zeitpunkt der Durchführung. Eine Datei, die beim Ingest in einer Kategorie bspw. 2a) eingruppiert wird, muss nicht die gleiche Einordnung bei einem späteren Prüfungslauf bekommen (z. B. bei regelmäßigen bestandserhaltenden Prozessen).

.....
 13 Vgl. Zur Kritik an JHOVE: "Ensuring long-term access: PDF validation with JHOVE?", verfügbar unter: JHOVE: https://www.cines.fr/wp-content/uploads/2015/01/PDF_A_JHOVE_Friese_28112014_en1.pdf [Abruf 4. 10. 2019] ebenso <https://lwn.net/Articles/688396/> [Abruf 4. 10. 2019] und <http://www.pdfa.org/2014/12/langzeitverfuegbarkeit-sichern-pdf-validierung-durch-jhove/?lang=de>

Kriterium	Kat. 0	Kat. 1	Kat. 2a	Kat. 2b	Kat. 3	Kat. 4
SIP ist virenfrei und konform zur DNS SIP-Spezifikation	Ablehnung	X	X	X	X	X
Fehlerfreie Dateiidentifikation (inkl. Codecs)			X	X	X	X
Alle Dateien unterstützter Formate im SIP sind nach Aussage eines Validators zu der jeweiligen Formatspezifikation konform. (Validation)			X		X	X
Alle Dateien unterstützter Formate können (falls notwendig) zu LZA-Format migriert werden. (LZA-Migrierbarkeit)				X	X	X
Alle Dateien im AIP sind im Sinne der DNS Spezifikation unterstützte Formate						X
Eignung zur dig. LZA in DNS	Keine, Neu-einl.	Bit-stream pres.	Teilw.	Voll.	Voll.	Voll.

Abb. 1: „Tabellarische Darstellung der Qualitätskategorien bei DNS“

Qualitätssystem bei DNS – Status Quo und Fazit

Die Qualitätskategorien wurden für DNS erfolgreich implementiert und werden nach ausgiebigen Test in Kürze in der produktiven Umgebung von DNS bereitgestellt.

Vorläufer der Qualitätskategorien war bisher ein Verfahren, welches die Übersteuerung eines „bekannten Fehlers“ ermöglichte. Dieses Verfahren wird zwar grundsätzlich beibehalten, wird aber durch die Qualitätskategorien entscheidend erweitert. Der signifikante Vorteil der neuen Kategorien ist es, dass sie bereits im Vorhinein nach eindeutigen Merkmalen möglicherweise auftretende Fehler spezifiziert. Dies hat überall dort Vorteile, wo hohe Datenmengen in automatischen Einlieferungsprozessen zu erwarten sind.

Ebenfalls wurde bei DNS eine Möglichkeit implementiert, Mindeststufen vom Einliefernden im so genannten „Contract“ (einer vom Einliefernden mitgelieferten Einstellungsdatei) mitzugeben. So kann für ein bereits erkanntes, problematisches Paket eine „Mindeststufe“ definiert werden, um es nach anderen Kriterien behandeln zu lassen. Auch so kann das digitale Archivsystem seinerseits Mindestvorgaben machen, die für alle Servicenehmer gelten. Über eine Webschnittstelle (REST-API) können Informationen zu den Kategorien eines Objekts ebenfalls abgefragt werden.

Damit entsteht für den Servicegeber und Servicenehmer ein transparentes Verfahren zur Nutzung bereit, welches auch bei hohen Datenmengen praktikabel einsetzbar sein wird.

Summary

The basic thesis of the article is that analogous to the processes of a classical archive, a digital archive archivable but damaged archived files can hardly reject. An analog document of high archival relevance with damage is, of course, taken over by an archive and professionally conservational treated and secured accordingly. In the practice of digital archiving using the example of the „DA NRW Software Suite (DNS)“, on the other hand, there is a dilemma for which the domain as a whole has to find solutions that are practicable even with the expected high data volume. Often, however, the assessment and the assessment of the severity of the damage is a complex issue in which there may be no good or even correct answer at second glance.

Author sketched a solution scenario, which assigns the individual steps of the ingest processing to certain quality categories, which build on each other in stages. Each level is characterized by a quality characteristic which expresses the suitability as a digital archive. These allow adequate storage and management even of problematic objects until, for example, a data archivist / data archaeologist or preservation manager can take care of the objects more precisely and purposefully. All assignments to quality categories are performed transparently and get a traceable entry in the long-term archiving metadata.

ERSCHLIESSUNG
UND
NUTZUNG

„ICH PACKE MEIN AIP UND NEHME MIT...“

Dateistrukturierung im Digitalen Magazin des Landesarchivs Thüringen (ThELMA)

Svenia Pohlkamp, Konrad Meckel

Das Projekt „Digitales Magazin des Freistaats Thüringen“ – Überblick und Stand

Das Projekt „Digitales Magazin des Freistaats Thüringen“ und Vorarbeiten dazu wurden bereits auf den Tagungen des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ in den Jahren 2013¹ und 2014² vorgestellt. Daher soll an dieser Stelle (wie auch im Vortrag, gehalten am 13.03.2019 in Prag) ein knapper Projektüberblick genügen: Das Projekt begann 2012 mit der Vorbereitung einer Ausschreibung zur Anschaffung einer OAIS-konformen Software. Als Ergebnis des Vergabeverfahrens erhielt die H&T Greenline GmbH Ende 2013 den Zuschlag. Damit begann die Umsetzungsphase des Projekts, die im Sommer 2018 in die Abnahmephase übergegangen ist. Seitdem werden letzte Arbeiten in der komplexen Software des Thüringischen Elektronischen Magazins (ThELMA) vorgenommen. Parallel werden Betriebskonzepte, wie IT-Sicherheitskonzept und Notfallkonzept, fertiggestellt. Außerdem wird die Infrastruktur für den Betrieb der Software in einem gesonderten Projekt, dem Integrationsprojekt, zusammen mit dem Thüringer Landesrechenzentrum (TLRZ) vorbereitet. Im nächsten Schritt ist die Inbetriebnahme von ThELMA mit Rollout und Schulungen vorgesehen, so dass das Projekt Ende 2020 abgeschlossen werden kann.

-
- 1 Filthaut, Jörg: Das Projekt Digitales Magazin des Freistaats Thüringen. Ein Werkstattbericht. In: Nolte, Burkhard; Huth, Karsten (Hg.): Standards, Neuentwicklungen und Erfahrungen aus der Praxis zur digitalen Archivierung. 17. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 13. und 14. März 2013 in Dresden. Halle/ Saale 2014. S. 81–88.
 - 2 Filthaut, Jörg: Einführung der digitalen Archivierung im Freistaat Thüringen. In: Filthaut, Jörg (Hg.): 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. 25–29.

Die Strukturierungskomponente

ThELMA soll zunächst drei Objektarten übernehmen und verarbeiten können: E-Akten, Daten aus Fachanwendungen sowie Ordner- und Dateistrukturen, sogenannte Dateisammlungen. E-Akten und Daten aus Fachanwendungen werden in ThELMA regelbasiert verarbeitet. Dateisammlungen besitzen im Gegensatz zu den anderen beiden Objektarten eine individuelle und beliebig verschachtelte Struktur. Zudem können die Unterlagen verschiedener Sachzusammenhänge in einer Dateisammlung vermischt vorliegen. Daher muss eine übernommene Dateisammlung nicht einem Archivpaket entsprechen. Eine manuelle Zuordnung bzw. Strukturierung der Ordner und Dateien zu Intellektuellen Einheiten ist in diesem Fall erforderlich:

Die vollständige Sichtung und Bewertung umfangreicher Ordner- und Dateistrukturen vor Ort in der abgebenden Stelle ist nicht praktikabel; umsetzbar ist lediglich eine grobe Bewertung. Für den Fall, dass übernommene Dateisammlungen Ordner bzw. Dateien enthalten, die nicht archivwürdig sind, sollen während der Bearbeitung der Übernahme im Ingestprozess Nachbewertungen möglich sein. Die gesamte Bearbeitung muss protokolliert werden, damit alle getroffenen Entscheidungen über Zuordnungen und Nachbewertungen zu einem späteren Zeitpunkt für abgebende Stellen, Archiv und Benutzer nachvollziehbar sind.

Die grundlegende Bearbeitung und Übernahme von digitalen Unterlagen wird im Landesarchiv Thüringen der fachlich zuständige Archivar durchführen, der von der Steuerungsstelle des Digitalen Magazins technisch unterstützt wird. Um die Akzeptanz des Verfahrens zu erhöhen, soll die Bearbeitung digitaler Unterlagen in einer Oberfläche erfolgen. Die Verwendung einer Sammlung von unterschiedlichen, vorgeschalteten Tools soll soweit möglich vermieden werden.

Die Bearbeitung einer Dateisammlung kann zeitaufwändig sein. Wünschenswert ist deshalb, die Bearbeitung unterbrechen und zu einem späteren Zeitpunkt fortsetzen zu können. Dazu muss jederzeit einfach erkennbar und nachvollziehbar sein, welchen Stand die Bearbeitung hat und welche Arbeiten noch erforderlich sind. Getroffene Entscheidungen und Metadateneingaben müssen bis zum vollständigen Abschluss der Bearbeitung geändert werden können.

Aufbau und Funktionen der Komponente

Mit ihrem Aufbau erlaubt die Strukturierungskomponente einen geordneten Zugang zu einer Ordner- und Dateistruktur in der Art eines Explorers, wie man ihn von der täglichen Arbeit am PC kennt. Anzeige und Zugang zur Struktur erfolgen in Spalten, da sich eine horizontale Darstellung der Struktur unter Berücksichtigung der technischen Gegebenheiten der Anwendung für die Bearbeitung als besser geeignet herausgestellt hat als eine vertikale Darstellung. Die Anzahl der Spalten (**Abb. 1**, Seite 96) kann nach den Vorlieben des jeweiligen Bearbeiters an die Anforderungen der Dateisammlung (bspw. Tiefe der Struktur) angepasst werden. Bei der Navigation in der Struktur unterstützt die Angabe des Pfades des jeweils ausgewählten Ordners bzw. der ausgewählten Datei innerhalb der übernommenen Dateisammlung. In Bezug auf den eigentlichen Zweck

der Komponente, nämlich die Strukturierung, wird sowohl das Bilden von Archivpaketen als auch die (nachträgliche) Kassation unterstützt. Beides erfolgt über einen Bereich rechts neben der Explorer-artigen Ansicht der Dateisammlung per Drag & Drop (**Abb. 1**): Dateien und Ordner, die kassiert bzw. archiviert werden sollen, können angeklickt und über den Bereich mit der gewünschten Funktion gezogen werden. Beim Loslassen der Maustaste wird die betreffende Funktion ausgeführt. Soll das gewählte Objekt archiviert werden, fragt die Anwendung unmittelbar bei Bildung des Paketes vordefinierte Metadaten zum neu anzulegenden Archivpaket vom Bearbeiter ab. Nach deren Eingabe ist das Paket innerhalb des Drag & Drop-Bereichs sichtbar (**Abb. 2**, Seite 97). Ihm können weitere Ordner und Dateien ebenfalls mit Hilfe des Drag & Drop-Verfahrens hinzugefügt werden. Alternativ können weitere Archivpakete angelegt werden, die dann ebenfalls im Drag & Drop-Bereich angezeigt werden.

Damit bei der Bildung der Archivpakete der Überblick erhalten bleibt und auch eine Unterbrechung der Arbeit möglich ist, wird in der Strukturierungskomponente mit einer farbigen und alphanumerischen Kennzeichnung gearbeitet. Bei Ordnern ist angegeben, wie viele der enthaltenen Ordner und Dateien noch nicht bearbeitet (gelb), einem Archivpaket zugeordnet (grün) oder nachträglich kassiert wurden (rot). Auch die Ordner und Dateien selbst sind mit Buchstaben entsprechend markiert („n“ = nicht bearbeitet, „a“ = einem Archivpaket zugeordnet, „k“ = zur Nachkassation vorgesehen). Siehe dazu das Beispiel „Brand HAAB“ in Abb. 2: Der Ordner enthält insgesamt 29 Ordner und Dateien, die alle einem gleichnamigen Archivpaket zugeordnet wurden. Das Archivpaket enthält inklusive des Ordners „Brand HAAB“ selbst insgesamt 30 Ordner und Dateien. Welche Ordner und Dateien zur Kassation oder zur Archivierung vorgemerkt sind, kann der Bearbeiter herausfinden, indem er im Drag & Drop-Bereich auf die jeweilige Schaltfläche bzw. die dort gelisteten Archivpakete klickt. Die enthaltenen Ordner und Dateien sind als Metadaten in einem Bereich unterhalb der Strukturierungskomponente aufgeführt (**Abb. 3**, Seite 98). Während der Bearbeitung wird der jeweils aktuelle Stand angezeigt und unterstützt so den Bearbeiter bei der Strukturierung.

Bis der Bearbeiter bewusst die Entscheidung trifft, dass er mit der Strukturierung fertig ist und diese Entscheidung durch Ausführen der Funktion „Bearbeitung abschließen“ bekräftigt, kann die erfolgte Bearbeitung überprüft und geändert werden. Wenn die Bearbeitung abgeschlossen wird, prüft die Anwendung, ob alle Ordner und Dateien bearbeitet wurden und ob für die Archivierung der Archivpakete alle notwendigen Informationen und Metadaten vorliegen. Sind Dateien und Ordner unbearbeitet geblieben, wird der Bearbeiter hierauf in einer Meldung hingewiesen. Der Bearbeiter kann entweder zur Strukturierungskomponente zurückkehren und die Bearbeitung fortsetzen oder durch Bestätigung der Meldung die automatische Kassation der Objekte auslösen. Die Archivierung bzw. Kassation der Unterlagen erfolgt im Nachgang in einem gesonderten Arbeitsschritt.

Resümee

Isabel Taylor hat 2016 Dateisammlungen als „hydraartige Matrjoschka“³ bezeichnet und damit die Herausforderung, die diese Objektart für die Archive mit sich bringt, bildstark zusammengefasst. Mittlerweile gibt es verschiedenste Anwendungen und Tools, mit denen die Archive dem „Monster“ Dateisammlung zu Leibe rücken. Zentraler Grundgedanke bei der Entwicklung der Strukturierungskomponente war, dass die Bearbeitung von Dateisammlungen aus Gründen des Akzeptanzmanagements und der Protokollierung integriert in einer OAIS-konformen Anwendung erfolgen sollte. Damit waren der Entwicklung der Komponente, genauso wie den restlichen individuellen Softwareanpassungen, die technischen Möglichkeiten und Grenzen der zugrundeliegenden Basissoftware gegeben. Alle Anforderungen an eine Komponente zur archivischen Bearbeitung von Dateisammlungen wurden in Zusammenarbeit mit dem Softwarehersteller in diesem Rahmen formuliert und umgesetzt. Dabei ist ein Stück Software entstanden, das die Navigation in einer Ordner- und Dateistruktur, die kleinteilige Strukturierung in Archivpakete und die (Nach-)Kassation ermöglicht. Das Landesarchiv ist erfreut, dass die Bearbeitung von Dateisammlungen ohne vorgelagerte Dritttools mit ThELMA möglich ist. Trotzdem bleibt das Grundproblem einer Dateisammlung bestehen: Je komplexer die Struktur, desto schwieriger ist es, den Überblick über die Struktur und die erfolgte Bearbeitung zu behalten. Zwar wurden auch dafür Hilfen entwickelt, ob diese aber genügen, um die alleinige Bearbeitung von Dateisammlungen mit der Komponente zu ermöglichen, wird wohl erst die Praxis zeigen. Das Projektteam ist unter anderem in diesem Punkt sehr gespannt auf die Rückmeldungen der Archivarinnen und Archivare, die ThELMA in ihrer täglichen Arbeit einsetzen werden. Daher gilt: Nach dem Projekt ist vor der Weiterentwicklung!

Summary

The project Digital Repository of the Freestate of Thuringia started in 2012 with the aim to establish a software solution for digital archiving in Thuringia. One of the main factors in choosing a solution was compliance to OAIS. Currently, the project is dealing with acceptance tests of the software ThELMA (Thüringisches Elektronisches Magazin) as well as finalization of concepts regarding security and technical operation. As of today, ThELMA will be launched at the end of 2020.

The paper is describing the component of the software solution ThELMA provided for structuring folder and file collections into archival information packages (AIP). The design of this component in collaboration with the developing

.....
 3 Taylor, Isabel: Eine hydraartige Matrjoschka. Wie wir die Fileablage eines staatlichen Schulamtes bewertet und erschlossen haben. Vortrag vom 29.02.2016 bei der 20. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ in Potsdam. URL: [https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2016/archivierung-von-unterlagen-mit-besonderen-strukturen/01_TAYLOR_Vortragsfolien%20\(29.02.16\).pdf](https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2016/archivierung-von-unterlagen-mit-besonderen-strukturen/01_TAYLOR_Vortragsfolien%20(29.02.16).pdf) [Abruf 09.05.2019].

company H&T Greenline was guided by the following factors and decisions: 1. to improve the acceptance of ThELMA among the archivists by avoiding a collection of third party tools for structuring during ingest, 2. to protocol the archivist's decisions structuring folders and files into AIPs or deleting them, 3. to enable the archivists to pause their structuring process, continue it at a later time and revise earlier decisions. The component allows the archivist to navigate through the folder and file structure, to assign objects to AIPs and to keep track of the process of structuring.

Zugang zur Struktur in Spalten,
Spaltenanzahl individuell anpassbar

→

Pfad

(Hervorgehoben durch einen roten Kreis)

→

Zeige 3 Spalten

(Hervorgehoben durch einen roten Kreis)

→

Hierher ziehen für neues Archivpaket.

Hierher ziehen zum Kassieren.

(Hervorgehoben durch einen roten Kreis)

Drag & Drop- Bereich für die Strukturierung in Archivpakete und die Nachkassation

Legende: nicht zugeordnet | archiviert | kassiert

Persistent-ID	d6e1c04b-2a0e-4e64-9868-17f5eda1f30c9
Erstellungszeitpunkt	08.03.2019 15:26:51
Bewertung	nicht bearbeitet
zugeordnet zu	Kein Archivpaket zugeordnet

Abb. 1: Oberflächenelemente der Strukturierungskomponente

Abb. 2: Strukturierung

Zeige 3 Spalten

/ filesystem-container(b0002816-790c-4764-a5f2-835a781755b4) / SB2 Kriminalitätsbekämpfung / Brand HAAB / Hinweis.txt

Komm3 Umweltdelikte	a 0 29 0	Brand HAAB	a 0 29 0	Einsatztagebuch	a 0 2 0
SB2 Kriminalitätsbekae...	n 76 30 0	Geldfälschung	n 3 0 0	Gutachten	a 0 3 0
SG1 Ordnungs- und Sch...	n 89 0 0	Innenministerkonferenz ...	n 28 0 0	Presse	a 0 7 0
		Quartalsberichte TIM	n 4 0 0	Sachstandsberichte	a 0 12 0
		Rauschgiftdelikte EF	n 6 0 0	Hiweis.txt	a
		Verdacht Diebstahl Schl...	n 30 0 0		

Hierher ziehen für neues Archivpaket.

Komm3 Umweltdelikte (27)

Brand HAAB (30)

Hierher ziehen zum Kassieren.

Zwei Archivpakete.
In Klammern steht die Anzahl der enthaltenen Ordner und Dateien.

Die farbige und alphanumerische Kennzeichnung macht deutlich, dass alle 29 Ordner und Dateien aus dem Ordner „Brand HAAB“ inkl. des Ordners selbst in das Archivpaket „Brand HAAB“ übernommen wurden.

Legende: nicht zugeordnet | archiviert | kassiert

Legende für die farbige und alphanumerische Kennzeichnung der Dateien und Ordner während der Strukturierung

Persistent-ID	3bbc5589-b032-47ef-a703-9ddc970270b1
Erstellungszeitpunkt	08.03.2019 15:26:53
Hashwert	29771fc440a4f10ff1b14fa6746dd1dbdb36883d095a2f28f65244ac770ab65
Hashverfahren	SHA-256
Bewertung	archivieren
Primärdatei	 Hiweis.txt
Größe	33 B
zugeordnet zu	Brand HAAB

Archivpaket

Erstellungszeitpunkt	08.05.2019 14:06:27										
Ersteller	pohlkamp(LATh)										
AIP-ID	0fbb646-732f-43e6-8202-5e6a42a8d1f0										
Technischer Titel	Brand HAAB										
Umfang	25										
Speichergröße	13,5 MB										
Speicher	<p><i>Primärdateien im ausgewählten Archivpaket</i></p> <table border="1"> <thead> <tr> <th>Name</th> <th>Beschreibung</th> <th>Standard</th> <th>Typ</th> <th>Persistenz-ID</th> </tr> </thead> <tbody> <tr> <td>Lokales Dateisystem</td> <td>Lokales Dateisystem für Testzwecke</td> <td><input checked="" type="checkbox"/></td> <td>FILESYSTEM</td> <td>c9e7c97f-8b66-454f-b588-fa887d8e1859</td> </tr> </tbody> </table> <p>1 Element : Anzeigen 5 Elemente</p>	Name	Beschreibung	Standard	Typ	Persistenz-ID	Lokales Dateisystem	Lokales Dateisystem für Testzwecke	<input checked="" type="checkbox"/>	FILESYSTEM	c9e7c97f-8b66-454f-b588-fa887d8e1859
Name	Beschreibung	Standard	Typ	Persistenz-ID							
Lokales Dateisystem	Lokales Dateisystem für Testzwecke	<input checked="" type="checkbox"/>	FILESYSTEM	c9e7c97f-8b66-454f-b588-fa887d8e1859							
Erhaltungsmaßnahmen											
Technische Hinweise											
Bereitstellungshinweise											
Elemente	<p><i>Ordner und Primärdateien im ausgewählten Archivpaket</i></p> <table border="1"> <thead> <tr> <th>Pfad</th> </tr> </thead> <tbody> <tr><td>/Brand HAAB</td></tr> <tr><td>/Brand HAAB/Einsatztagebuch</td></tr> <tr><td>/Brand HAAB/Einsatztagebuch/lagebuchHAAB.docx</td></tr> <tr><td>/Brand HAAB/Einsatztagebuch/lagebuchHAAB.pdf</td></tr> <tr><td>/Brand HAAB/Gutachten</td></tr> <tr><td>/Brand HAAB/Gutachten/Anl.1.pdf</td></tr> <tr><td>/Brand HAAB/Gutachten/Anlage2.pdf</td></tr> </tbody> </table> <p>30 Elemente : Anzeigen 7 Elemente</p>	Pfad	/Brand HAAB	/Brand HAAB/Einsatztagebuch	/Brand HAAB/Einsatztagebuch/lagebuchHAAB.docx	/Brand HAAB/Einsatztagebuch/lagebuchHAAB.pdf	/Brand HAAB/Gutachten	/Brand HAAB/Gutachten/Anl.1.pdf	/Brand HAAB/Gutachten/Anlage2.pdf		
Pfad											
/Brand HAAB											
/Brand HAAB/Einsatztagebuch											
/Brand HAAB/Einsatztagebuch/lagebuchHAAB.docx											
/Brand HAAB/Einsatztagebuch/lagebuchHAAB.pdf											
/Brand HAAB/Gutachten											
/Brand HAAB/Gutachten/Anl.1.pdf											
/Brand HAAB/Gutachten/Anlage2.pdf											

Abb. 3: Inhalt des Archivpakets

VON DER ARCHIVVORSTUFE ZUM ARCHIV

(Forschungs-)Datenmanagement im Kontext von Gestaltung und Kunst

Jürgen Enge, Tabea Lurk

Einleitung

Das digitale Archiv der Hochschule für Gestaltung und Kunst (HGK) FHNW Basel wird seit 2017 gemeinsam vom Center for Digital Matter (DIGMA) und der Mediathek aufgebaut. Es beabsichtigt, die Geschichte der Hochschule, ihre digitalen Lehr- und Forschungsmittel sowie die ihrer affilierten Kontexte dokumentarisch zu erschliessen, dauerhaft aufzubewahren und gemeinsam mit anderen Wissensressourcen (z. B. Bibliotheksressourcen) durchsuchbar zu machen. Auch wenn die Hochschule von überschaubarer Grösse ist, zeichnet sie sich durch eine bemerkenswerte Dynamik und Agilität aus, deren Wesenhaftigkeit in den entsprechenden Infrastrukturen reflektiert werden sollte. Am deutlichsten wird dies vielleicht in der Art und Weise, wie mit Datenlieferant*innen und ihren Beständen umgegangen wird.

Zudem kann das digitale Archiv der HGK als strategisches Instrument begriffen werden, das die Hochschule und ihre Angehörigen auf sanfte Weise an die Digitalisierung heranzuführt: Die Übernahmeprozesse gestalten sich stets dialogisch; neue Werkzeuge, digital motivierte Gedankengänge und Abläufe werden vermittelt, die künftig frühzeitig(er) berücksichtigt werden oder sogar gemeinsame Entwicklungsprojekte anstossen.

Die Aufnahme der Konvolute und Sammlungen erfolgt stets fallbasiert, auf Initiative von (Einzel-)Personen und nicht aufgrund eines Erlasses. Mit Blick auf die Aktenbildung steht uns dieser interessensgeleitete Weg offen, da der Grossteil der aufbewahrten Inhalte keine obligationsrechtlichen Zwecke bedient: Daten, welche aufgrund gesetzlicher Aufbewahrungsvorschriften archiviert werden müssen, werden vom Stab bzw. dem FHNW Präsidium aufbewahrt oder den jeweiligen Stellen zur Übernahme angeboten.¹ Die gesetzliche Revisionsfrist von Abschlussarbeiten beträgt lediglich 10 Jahre und wird an der HGK bisher von den Instituten selbst gemanagt, da im Falle der Revision neben den von uns dokumentierten Thesen und Ressourcen zumeist auch praktische Lehr- und Lernresultate des jeweiligen Abschlusses zur Disposition stehen. Dass die medial überlieferten Inhalte der HGK hochschulintern dennoch

.....
1 Unsere Bestände enthalten z. B. keine Studierendendaten oder Studieninformationen wie sie die Fachgruppe 8 „Archive der Hochschulen und wissenschaftlichen Institutionen“ im „Verband deutscher Archivarinnen und Archivare“ (VdA) behandelt (vgl. Fussnote 8).

ausgesprochen wichtig erscheinen und kulturelle, bildungspolitische oder strategische Zwecke erfüllen, steht auf einem anderen Blatt.² Aber genau hier, an den semantisch wertvollen Inhalten, die fakultativ gesammelt werden können, setzt das digitale Archiv der HGK an³. Entsprechend folgt die Aufnahme von Inhalten einfachen Prämissen:

- *Institutionelle Affiliation*: Die angebotenen Inhalte müssen an der HGK entstanden, von ihr/Hochschulangehörigen in Auftrag gegeben oder nachweislich mit Personen, Instituten oder Themen der HGK assoziiert sein, sodass ihre Aufbewahrung der Forschung und Lehre dient und Bezüge benannt werden können.⁴
- *Geklärte Rechtelage*: Die einfachen Nutzungsrechte an den Inhalten müssen der HGK übertragen werden. Inhalte, die dauerhaft für alle Nutzungsgruppen unzugänglich bleiben, finden nur in absoluten Ausnahmefällen Einzug ins digitale Archiv.⁵ Kontrollierte Zugangsformen sind hingegen selbstverständlich möglich.
- *Ökonomische Vertretbarkeit*: Es fallen keine Sonder- oder Akquisitionskosten an, die jenseits der Aufwände für die Erschließung, Vereinnahmung, (dauerhafte) Archivierung in der Mediathek zu Buche schlagen.
- *Kulturelle, historische, regionale, strategische Relevanz*: Die Inhalte sind nachweislich von besonderem kulturellem, historischem oder regionalem Wert – sie tragen also zur Entwicklung und Bedeutung der HGK bei und ermöglichen Formen der Vernetzung, die der strategischen Ausrichtung der HGK dienen.⁶

Bestandsübersicht

Ein Blick auf die Bestände verdeutlicht die Heterogenität der Inhalte, wobei ein Bestand als «Niederschlag einer schöpferischen, geschäftlichen oder anderweitigen Tätigkeit» verstanden wird, wie ihn Philipp Messner aus der institutionellen Praxis der Archivtheorie ableitet.⁷ Neben der Provenienz spielt die Sacherschließung (Pertinenzprinzip) eine wichtige Rolle.

-
- 2 Die an Kunsthochschulen entstehenden Arbeiten gewinnen nicht nur bei später erfolgreichen Künstler*innen und Gestalter*innen an Wert, sondern sind insgesamt von historischer Relevanz. (vgl. z. B. die Lehr- und Vermittlungsforschung)
 - 3 Vgl. Baumann, Laurent; Messerli, Franco; Mumenthaler, Samuel: *Memoriav Bulletin*: Zugang. Bern 2015. Online: https://memoriav.ch/wp-content/uploads/2015/10/Bulletin-22_Web.pdf [Abruf 4. 10. 2019].
 - 4 Dies trägt der Forschungsrelevanz von künstlerischen resp. gestalterischen Tätigkeiten Rechnung, wie sie in den «Strategischen Leitlinien FHNW 2025» (2019, S. 12) niedergeschrieben sind.
 - 5 Ein solcher Bestand sind z. B. Thesen oder Abschlussarbeiten, die auch dann (sofern möglich) vollständig aufbewahrt werden, wenn die Institute sie aufgrund von mangelnder Qualität nicht verbreitet/publizieren möchten.
 - 6 Jahresrückblick FHNW. Lurk, Tabea (2019): Ankommen und Zusammenwachsen. In: Langkilde, Kirsten; Hochschule für Gestaltung und Kunst FHNW (Hrsg.) *Imagine* (Jahresbericht 2018). Basel: Peter Median Verlag, S. 261-270.
 - 7 Messner, Philipp: *Eine andere archivische Praxis*, Brand-New-Life – Magazin für Kunstkritik. Zürich 2017. Online: <http://brand-new-life.org/b-n-l-de/eine-andere-archivische-praxis/> [Abruf 4. 10. 2019]. .

Unter den hochschuleigenen, digitalen Sammlungen sind zu erwähnen:

- die *Videosammlung des Instituts Kunst*, die der Künstler, Kurator, Kunsttheoretiker und ehemalige Leiter des Instituts Kunst, René Pulfer, seit den 1990er Jahren bis 2014 aufgebaut hat, bis die HGK auf das Dreispitz Areal gezogen ist und die Mediathek neu organisiert wurde. Hierbei handelt es sich um ca. 2500 diskbasierte Datenträger mit vornehmlich Film- und Videoarbeiten sowie Referenzbeispielen der Videokunst. Die Sammlung wird bis heute, wenn auch mit unterschiedlichem Engagement und Strategien, weitergeführt.
- das Archiv des *Instituts Industrial Design* (2006-2015), das projektbasierte Arbeiten und Entwürfe dokumentiert und den wiederkehrend stattfindenden Modulen zuordnet. Die sowohl formal (Verknüpfung von Studienverlauf, Lehrangeboten und Studierendendaten)⁸ als auch inhaltlich gepflegte Struktur klassifiziert die Inhalte entlang einem hierarchisch gegliederten Thesaurus und wurde aufgrund der hochwertigen Datenstruktur zum Vorbild für diverse Bestands-Modellierungen.
- Teile der *Thesenarchive der Institute Kunst* (1995/2012-2018), *Hyperwerk* (2002-2016, ohne 2007) und *Masterstudio* (2010-2019).
- seit 2017 die *DiplomHGK-* bzw. *NextGeneration-Website*, welche kurze Abstracts, Bild-, Audio-, Video- und Textmaterial aller Absolvent*innen eines Jahrgangs der HGK präsentiert und die gleichnamige Jahresausstellung begleitet.
- seit Mitte 2018 die Inhalte der *Campus-der-Künste-Website*, die als eine zentrale Plattform der Aussenkommunikation funktioniert und wichtige Veranstaltungen, News etc. ankündigt.⁹
- der Teilbestand *ACT Basel* (2004-2018), des gleichnamigen, schweizweit ausgetragenen Performance-Festivals ACT. Aussergewöhnlich ist dieses Konvolut, weil das Festival hochschulübergreifend ausgetragen wird und die Teilnehmenden aller Hochschulen an allen Orten ihre Performances auführen können. Das Konvolute enthält über 900 Ansetzungen von über 340 Künstler*innen und/oder Künstlergruppen mit etwas mehr als 14.000 Medien (Video-, Bild-, Ton- und Textmaterialien).
- seit 2019 fotografisches Dokumentationsmaterial, das Hans-Peter Huser in seiner Arbeit für die HGK Kommunikation zwischen 2007 und 2018 fotografiert oder in Auftrag gegeben hat und alle wichtigen Anlässe der Hochschule in diesem Zeitraum dokumentiert.

Zu den Inhalten aus Forschungsprojekten und den forschungsrelevanten Sammlungen gehören:

- Resultate des Forschungsprojekts *To Perform* (2012 ff.), das sich mit der Aufführungspraxis der Neuen Musik beschäftigte und federführend von der Hochschule für Musik FHNW realisiert wurde. Im Rahmen von fünf Fallstudien wurden die Komponisten Vinko Globokar, Dieter Schnebel, Heinz Holliger, Christoph Caskel, Peter Maxwell Davies und Péter Eötvös bei der Erarbeitung einer Aufführung (inkl. Instrumentenbau, Spielanleitungen etc.)

8 Zur Relevanz dieser Strukturverknüpfung vgl. Nippert, Klaus: Archivisches Dokumentationsprofil für Studierendendaten aus elektronischen Fachverfahren. In *Archivar* 2016: 69 (3) S. 252.

9 Das entlastet längerfristig die Erschliessungsarbeit und hilft, versehentliche Lücken zu vermeiden.

dokumentiert, um mithilfe der Notationen, aber eben auch von Videoaufzeichnungen, Interviews, Proben- und Bildmaterial spätere Wiederaufführungen zu ermöglichen. Die Videodokumentation erfolgte durch Angehörige des Instituts Kunst unter der Leitung von Reinhard Manz.¹⁰

- Die Online-Publikation eines kuratierten Datenauszugs des Forschungsprojektes *Grenzgang - Künstlerische Untersuchungen zur Wahrnehmung und Vermittlung von Raum im trinationalen Grenzgebiet* (2014-2015), das vom Schweizerischen Nationalfond gefördert wurde und am Institut Lehrberufe für Gestaltung und Kunst realisiert wurde.
- die Resultate der *Videowochen im Wenkenpark* (1984, 1986, 1988),¹¹ einem frühen Videokunstkurs, der von René Pulfer (kuratorisches Konzept) und Reinhard Manz (Produktion) in der Nähe von Basel organisiert wurde und an dem heute weltbekannte Videokünstler*innen aus unterschiedlichen Ländern und der Schweiz teilgenommen und währenddessen Videoerwerke produziert haben.

Unter den in Kooperation aufgebauten, digitalen Sammlungen finden sich

- jene Video- (und Film-)arbeiten, die nach dem Ausstellungsprojekt Σ *Summe 2017* eine Berechtigung zum Verbleib in der Mediathek erhielten.
- die an der Musikakademie Basel FHNW zwischen 1974 und 2017 aufgebaute *Basler Bibliographie zur Historischen Musikpraxis*, die als ausgesprochen hochwertige Quellensammlung 2017 eingestellt wurde und die als historisches Dokument nun zur weiteren Bearbeitung z. B. via Zotero zugänglich ist.¹²
- Teile der *Digitalen See*, einem massgeblich in Basel verorteten digitalen Sammlungsbestand, der vornehmlich videografische Videodokumentationen zur Performancekunst, verzeichnet und aufbewahrt. Wie die Betreiberinnen Muda Mathis, Videokünstlerin und ehemalige Dozentin im Institut Kunst, Sus Zwick und Chris Regn schreiben, ist die *Digitale See* «kein Ort, sondern eine Idee und Haltung», die «zum Ziel [hat], Kunst, in diesem Fall Performance Dokumentationen, der freien ungebundenen Szene im Internet zugänglich zu machen».¹³ Aus diesem wachsenden Bestand sind als abgeschlossene Teilbestände erschlossen: die *Performance Chronik Basel 1987–2006*,¹⁴ das *Feministische* Improvisatorium* (2018)¹⁵ und die sogenannten *Radiosendungen* (2018)¹⁶ des Kasko. Aktuell läuft ferner die Integration eines Grossteils der videografischen Dokumentationssammlung des Kasko - Kaskadenkondensator: Raum für aktuelle Kunst und Performance (1994–2019).
- die Podcasts des Kulturprojekts *Digital Brainstorming* des Migros Kulturprozent, die der Journalist, Autor, Medien- und Kulturschaffende Do-

10 To Perform: https://www.fhnw.ch/de/forschung-und-dienstleistungen/musik/hochschule-fuer-musik/projekte/to_perform [Abruf 4. 10. 2019]. Die Archivierung des Projekts ist derzeit sistiert.

11 Wenkenpark: <http://mediathek.hgk.fhnw.ch/wenkenpark.php> [Abruf 4. 10. 2019].

12 Basler Bibliographie zur Historischen Musikpraxis: https://www.zotero.org/groups/2171465/basle_bibliography_for_historical_performance_practice [Abruf 4. 11. 2019].

13 Performance Chronik Basel: <https://mediathek.hgk.fhnw.ch/pcb.php> [Abruf 4. 10. 2019].

14 Ebd.

15 Feministisches Improvisatorium: http://mediathek.hgk.fhnw.ch/feministisches_impro.php [Abruf 4. 10. 2019].

16 Radiosendungen: <http://mediathek.hgk.fhnw.ch/radiosendungen.php> [Abruf 4. 10. 2019].

minik Landwehr zwischen 1998 und 2019 realisiert hat und prominente Vertreter*innen der aktuellen Szene der Medienkultur und -kunst zu Wort kommen lässt.

Zu den wichtigsten externen Quellen des Katalogsystems, deren Metadaten in den SOLR-Volltextindex zwar aufgenommen und dort aktualisiert, nicht aber dauerhaft gespeichert werden, gehören

- die bibliographischen (Meta-)Daten der *Swissbib* (inklusive Nationallizenzen), der Open Access Repositorien für Bücher (DOAB) und Zeitschriften (DOAJ), Springer und DeGruyter Open, die lizenzierten Zeitschriften der HGK, der FHNW und von Swissuniversites, die in der elektronischen Zeitschriftendatenbank EZB verwaltet werden, die Nationallizenzen und weitere, frei zugängliche, relevante Sammlungen z. B. der ETH Zürich.
- die Metadatensammlungen von *videokunst.ch* und von *EuropeanArtNet*, wobei letztgenanntes vornehmlich Künstler*innen mit Referenzen auf Ausstellungen und Rezensionen benennt.¹⁷

Auch wenn die Teilbestände hier gemäss der zeitlichen Abfolge ihrer Erschliessung aufgelistet wurden, greift eine chronologische Betrachtung der Bestandsgeschichte zu kurz. Zum einen wurden in allen vier Bereichen gleichzeitig gesammelt. Zum andern wurde auf eine systematische Erschliessung entlang der dossiers-bildenden Bestandscluster, also z. B. zuerst alle Abschlussarbeiten/Thesen etc., verzichtet, um die unterschiedlichen Mentalitäten und Geschwindigkeiten der Institute zu respektieren. So konnten individuelle Übergabeformen gefunden werden. Während einige Institute derzeit noch entspannt beobachten, wie die Prozesse ablaufen, welche Aufwände effektiv beim Zusammenstellen der Daten anfallen oder wie sich die relativ einfachen, stets niederschweligen Workflows gegen die Entlastung der institutsinternen Aufbewahrung und der Regelung von Zugriffskonventionen z. B. bei der Herausgabe von Thesen rechnen, haben andere die Chancen der Zusammenarbeit mit dem Archiv erkannt und die Übergabe z. B. der Thesen direkt in die semesterbasierten Abläufe ihrer Sekretariate integriert. Wieder andere waren experimentell unterwegs und haben die Gelegenheit der strukturierten und dauerhaften Aufbewahrung genutzt, um Aspekte ihrer Archivkonzepte umzusetzen oder eigene Archivierungshaltungen zu etablieren, die in künftigen Projekten von Anfang an berücksichtigt werden. Nicht zuletzt gab es Einzelfälle, in denen produktbasiert aufgrund des technischen Wandels (System- und Software-Obsoleszenz) die Nähe zum Archiv gesucht wurde.¹⁸

Aus- und Weiterentwicklung

Da längst nicht alle archivrelevanten, hochschuleigenen und strategisch spannenden Konvolute erschlossen sind, wächst das digitale Archiv in unter-

17 Derzeit wird geprüft, ob die Integration der Metadaten von Europeana Collections auch möglich ist, da diese Inhalte die Suchbedürfnisse unserer Nutzer*innen sehr gut befriedigen und den Reflexionsradius erweitern könnten.

18 Eine klassische Notsicherung besteht in der Emulation oder Virtualisierung einer Out-of-Service gehenden Datenbankinfrastruktur, wie sie z.B. das Institut Industrial Design seit (2017) auf der Basis einer Filemaker-Applikation nutzte. Parallel zur mittelfristigen Verstetigung in der virtuellen Maschine (VM) wurden die Daten auch ins Katalogsystem der Mediathek und das digitale Archiv der HGK testweise migriert.

schiedliche Richtungen. Die kontinuierliche (Weiter-)Entwicklung folgt einem modularen Ansatz, der sich sowohl dynamisch an neue Softwarearchitekturen, Werkzeuge, Prozesse oder aktuell(st)e technologische Entwicklungen anforderungsbasiert anpasst, als auch die wandelbaren hochschulpolitischen oder strategischen Zielsetzungen umsetzen kann, ohne dabei die Dauerhaftigkeit der Inhalte zu gefährden. Diese grundlegende Flexibilität erscheint aufgrund der aktuellen Rahmenbedingungen unabdingbar und durchzieht sowohl die administrativ-strategische als auch die technologische Ausrichtung des Archivs.

Auf die Bestandsebene heruntergebrochen, bedeutet das beispielsweise, dass die Frage, wie repräsentativ die ggf. neue Struktur einer Sammlung für die bestehenden oder künftige Konvolute ist, darüber entscheidet, ob die zu übernehmenden Daten auf die bestehenden Datenmodelle gemappt oder eine neue Datenstruktur integriert wird. Gleiches gilt für die Bearbeitungstools: so wurde die Erkennungskaskade des Werkzeugs *Indexer* (s. u.) Ende 2018 um die Softwarebibliothek Siegfried¹⁹ erweitert, weil erstmals grössere Mengen an Blu-Ray-Videodaten ins Archiv gelangten, die von den bestehenden Werkzeugen nicht korrekt erkannt wurden. 2019 wurde die *metadata-extractor*-Bibliothek eingebunden,²⁰ um Metadatenformate wie z. B. Exif, IPTC, XMP, JFIF und andere aus digitalen Fotografien extrahieren zu können.²¹

Anforderungen und Zielsetzung

Eine der wichtigsten Ressourcen des digitalen Archivs der HGK sind die Datenlieferant*innen. Sie haben die Inhalte über einen längeren Zeitraum produziert oder kumulierend zusammengetragen und kennen die Hintergründe, Rechtelage und Kontexte. Sie haben in der Regel individuelle Ordnungen oder Organisationsformen entwickelt, von denen aus inhaltlichen (Authentizität, Wahrung von Entstehungs- und Überlieferungskontexten) aber auch arbeitsökonomischen Gründen so viel wie möglich ins digitale Archiv der HGK übernommen werden soll.

Das Anforderungsportfolio sollte ferner folgende Wünsche berücksichtigen:

- *Ein System / eine Software für unterschiedliche Datentypen:* Trotz der fachlichen, thematischen und medialen Spannbreite sowie der variierenden Erschließungstiefe sollten Katalogisierung und Aufbereitung über ein einziges Werkzeug (Frontend) funktionieren und sich quasi out-of-the-box den verschiedenen Bedürfnissen und Denktraditionen anpassen. Damit können ferner die Datenlieferant*innen ggf. direkt in die Erschließung oder Erfassungskontrolle eingebunden werden.
- *System- und Ortsunabhängigkeit:* Das Erschließungssystem sollte systemunabhängig (Mac, PC, Linux) einsetzbar sein und das Editieren der Inhalte auch dann erlauben, wenn die Datenlieferant*innen nicht (mehr) der Hochschule affiliert oder vor Ort sind.

19 <https://www.itforarchivists.com/siegfried/> [Abruf 4. 10. 2019].

20 <https://github.com/drewnoakes/metadata-extractor> [Abruf 4. 10. 2019].

21 Speziell hat uns die firmenspezifische MakerNote (=Kamera-ID) interessiert, welche Fotografien häufig eindeutig identifiziert und bei der Übernahme der Dokumentationsfotos der HGK Kommunikation anfiel.

- *Dauerhafter Zugriff*: Neben der Option des verteilten Arbeitens sollten die Datenlieferant*innen ggf. auch nach der Übergabe die Möglichkeit haben, auf die strukturiert aufbereiteten Inhalte ihrer Sammlungen – jenseits des Recherche/Wiedergabe Interfaces im Internet – zuzugreifen.
- *Open Access*: Potentiell sollte die Möglichkeit zur Nachnutzung und Erweiterung der Archivkonvolute eingeräumt werden und zwar auf der Ebene der (Meta-)Daten. Wenn die Dateneigner*innen dies wünschen, wie beispielsweise die Basler Bibliographie (s. o.). Gerade mit Blick auf die Archivierung von Forschungsdaten ist dies ein durchaus relevanter Aspekt.

Neben der niederschweligen Datenübergabe und (anhaltenden) Datenpflege sprachen eine ganze Reihe an Gründen für die Nutzung von Zotero.²² Die open source Software wurde von der Corporation for Digital Scholarship und dem Roy Rosenzweig Center for History and New Media für die Verwaltung von (wissenschaftlicher) Literatur entwickelt und unterstützt mittlerweile 34 unterschiedliche Medientypen, darunter die Standard Archivformate. In der Mediathek spielt Zotero als Zwischensystem in der Archivvorstufe, bei der Aggregation und bei der fließenden Erschließung eine zentrale Rolle.

Zu den übergeordneten Zielen gehörten:

- *Rasche Sichtbarkeit der Resultate*: Da unsere Datenlieferant*innen in der Regel noch leben, möchten sie ihre Inhalte möglichst schnell online sehen. Es braucht also einfache Workflows, welche die im Zotero verwalteten Inhalte leicht ins Rechercheinterface des Integrierten Katalogs ausspielen.
- *Offene Schnittstellen*: Damit die hauseigene Software in den Vereinnahmungs- und Wiedergabeprozessen auf die Daten in Zwischensystemen zugreifen kann, sind offene, dokumentierte Schnittstellen gefordert.
- *Statistische Auswertung*: Optional sollen statistische Auswertungsmechanismen bedient werden können, wobei hier das technologisch Machbare von der datenschutztechnischen Legalität reguliert wird.

Workflows

Aus dem Anforderungsprofil haben sich ferner die Workflows ergeben, deren Ausgestaltung an die Quelle angepasst und bei Bedarf optimiert wird. Bestimmend sind hierbei Aspekte wie:

- *Materiallage*: Welche Daten liegen in welcher Form vor, woher stammen sie und wie sollen sie später zugänglich sein?
- *Metadatenlage*: Welche Beschreibungsdaten oder Kontextinformationen sind wie greifbar?
- *Anvisierte Erschliessungsprozesse und -tiefe*: Wer soll die Inhalte erschliessen und welche Vorkenntnisse liegen vor?
- *Wiedergabezwecke/Nutzungsform*: Für welche Zielgruppe sollen die Inhalte erschlossen werden und für wen sollen sie später zugänglich sein?

Mit Blick auf die Verarbeitungsworkflows lassen sich die Schritte der Archivvorstufe, der Erschließung, der Medienverarbeitung/Zusammenführung,

.....
 22 Corporation for Digital Scholarship; Roy Rosenzweig Center for History and New Media: Zotero Documentation [de:start]. Massachusetts 2019. Online: <https://www.zotero.org/support/de/start> [Abruf 4. 10. 2019].

der Wiedergabe sowie schliesslich der digitalen Archivierung unterschieden (vgl. Abbildung 1, unterste Zeile, Seite 114).

Archivvorstufe

Auf die Anfrage zur Übergabe oder Übernahme von Beständen folgt die *Sichtung* des Bestands.²³ Dabei werden neben rechtlichen Aspekten die Prozesse und Erschliessungsworkflows definiert (vgl. Abbildung 1). Bei der anschliessenden *Datenlieferung* werden die Daten – zumeist mittels Memorysticks, Hard-Disks oder in Einzelfällen gerippt – auf die FHNW-Infrastruktur überführt (kopiert) oder über die Serverlaufwerke direkt freigegeben.²⁴

Es folgt die inhaltliche *Erschliessung*, die fast immer *Zotero* nutzt. Daher werden die Quelldaten häufig in einem mehrstufigen Verfahren für dieses System vorbereitet. Unterschieden wird dann zwischen der Aufbereitung von Daten für den *Zotero*-Workflow und den sog. *Indexer*-Workflow²⁵, der ebenfalls ins *Zotero* mündet, aber keine vorherigen, inhaltlichen Aufbereitungsleistungen erfordert.

Im Unterschied zu vollautomatisierten Ingestprozessen wird die Vollständigkeit der übergebenen Daten ebenso wie die Metadatenqualität händisch geprüft. Weil zumeist diverse kleinere Anpassungen durchgeführt werden müssen, sprechen wir von *Rohdaten*.

Aufbereitung von (Roh-)Daten

Je nach Quelle werden die Metadaten teilweise oder vollständig durch die Mediathek erstellt. Sofern Metadaten vorhanden sind, korrigiert/harmonisiert die Mediathek diese.

Der Zugang zu den jeweiligen Daten und die Einbettung der Datenbasis variiert zwischen:

- *Files im Dateisystem*: Die einzelnen Dateien liegen in strukturierten Foldern vor, sodass die Kontexte, die Erschliessungssystematik, häufig Ortsangaben etc. ersichtlich sind. Gerade bei Daten, die wiederkehrende Metadateninformationen haben, wie z. B. bei Abschlussarbeiten der Dokumenttyp, Sammlungszugehörigkeit (Institut/Fachbereich, Jahrgang), bietet sich ein Zwischenschritt via Excel an, in dem die Datensätze verzeichnet werden (vgl. Abbildung 2 und 4, Seite 115, 117).²⁶
- *Daten plus Tabelle*: Liegen Metadaten-Informationen bereits in Tabellenform (z. B. Excel, CSV) vor, muss lediglich die Konsistenz der Einträge überprüft,

23 Während bei der Übergabe die Datenlieferanten das Angebot offerieren, fragt die Mediathek bei der Übernahme die Dateninhaber aktiv an.

24 Die Speicherumgebung der FHNW besteht derzeit aus einem NetApp-Speicher. Die Datensicherung erfolgt als Block-Backup von Snapshots (tägliches inkrementelles Backup (5/7) und monatliches Full-Backup). Das Backup wird hybrid auf Tape und Disk gesichert

25 Die Kernfunktionalität des Indexers ist die Erstellung eines umfassenden Indexes, also «eines Hilfsmittels zur Auffindung von Sachbetreffen, Personenbetreffen oder Toponymen» Westphal, Sina: Indizieren. In: Terminologie der Archivwissenschaft, Archivschule Marbach. 2019. Online: <http://www.archivschule.de/uploads/Forschung/ArchivwissenschaftlicheTerminologie/Terminologie.html> [Abruf 4. 10. 2019]. Zur Funktionalität des Indexers vgl. Enge, Jürgen; Kramski, Heinz Werner: Exploring Friedrich Kittler's Digital Legacy on Different Levels: Tools to Equip the Future Archivist. In IRPES 2016, Bern 2016, S. 229–236.

26 In vielen Datenbank-Systemen werden derartige Inhalte über (Satz-)Schablonen erfasst.

ggf. angepasst und das Endresultat so auf die Ziel-Schemata von Zotero gemappt werden, dass der Import möglich wird.

- *Datenbank*: Beim Export von Metadaten aus Datenbanken besteht die Herausforderung darin, alle Inhalte bzw. Tabellen und die ursprünglichen Verknüpfungen sauber abzubilden. Bisher konnten alle Inhalte direkt abgebildet werden; sollten dies nicht mehr möglich sein, weil die Einträge nur als Ergebnisse (z. B. einer komplexen Suche aus einem Datenbanksystem) vorliegen, wird vermutlich das *Performance*-Prinzip angewandt, das Harvey und Weatherburn mit Blick auf prozessbezogene Unterlagen als weit verbreitet und anerkannt beschrieben haben.²⁷ Das anschließende Mapping erfolgt analog zum Vorgehen bei Tabellen.
- *(inkludierte) Objektmetadaten*: Die in Einzelmedien vorhandenen, semantischen und technischen Metadaten können z. T. automatisiert erkannt und extrahiert werden. Bisher findet neben den im *Indexer*-Workflow hinterlegten Routinen primär die Extraktion fotografischer Bildinformationen mittels *metadata-extractor* Anwendung.

Indexer-Workflow

Wenn viele Dubletten, unterschiedliche auch historisch variierende Medienformate und eine erhöhte Dauerhaftigkeit erwartet werden, nutzen wir den *Indexer*-Workflow. Der *Indexer* extrahiert jene Objektmetadaten, die im OAIS-Modell als Repräsentationsmetadaten bezeichnet werden.²⁸ Dabei durchlaufen die Daten nach dem Kopieren auf die hauseigenen Systeme folgende Schritte:

- *Indexierung*: Die vereinnahmten Daten werden automatisiert indexiert. Zunächst werden alle Dateien in einer Index-Datenbank erfasst, die auch den Dateinamen und Speicherort dokumentiert.²⁹ Die Erkennungskaskade erstellt dann mithilfe diverser Softwarebibliotheken technische Metadaten (Dateiformationen), generiert Volltexte sowie bei bekannten Bild- und Video-Formaten Vorschaubilder und sie erkennt Dubletten (auf der Basis von Checksummen). Dabei finden u. A. folgende Softwarebibliotheken Anwendung: Apache Tika; ffmpeg; gvfs-info; imageMagick; Libmagic; OpenLayers; Siegfried; Sonogram (selbstentwickeltes C++ Programm). Auch wenn die Daten damit noch lang nicht semantisch erschlossen sind, werden wichtige Informationen sichtbar und können für die weitere Bearbeitung genutzt werden.
- *Inventarisierung*: Mithilfe von selbst-generierten Inventarnummern werden die *relevanten* Inhalte aus dem Datenpool des *Indexers* gefischt und die Informationsobjekte geschnürt.³⁰ Dabei wird zwischen Einzel-Inventarnummern (ein Objekt = eine Inventarnummer) und Gruppen (n-Objekte bekommen als Bandel je die gleiche Inventarnummer; die Files werden hochge-

27 Harvey, Ross; Weatherburn, Jaye: Preserving Digital Materials. Rowman & Littlefield 2018, S. 87.

28 Vgl. nector-Arbeitsgruppe OAIS-Übersetzung / Terminologie: Referenzmodell für ein Offenes Archiv-Informationssystem – Deutsche Übersetzung, Version 2, Frankfurt am Main 2013, S. 60, urn:nbn:de:0008-2013082706. <https://d-nb.info/104761314X/34> [Abruf 4. 10. 2019].

29 Dateinamen und Speicherorte geben wichtige Hinweise auf den Kontext und die ursprüngliche Ablagesystematik, die häufig für die Dossierbildung hilfreich ist.

30 Ein Informationsobjekt kann mehrere Dateien (Datenobjekte) enthalten und wird durch Repräsentationsinformationen erweitert. Vgl. nector 2013, S. 19.

zählt) unterschieden (vgl. Abbildung 3, Seite 116). Pro Inventarnummer wird ein Handle vergeben.³¹

Zotero-Import

Nach der Aufbereitung der (*Roh*-)Daten werden die Metadaten ins Zotero übertragen oder erscheinen dort, im Falle des *Indexer*-Workflows, vollautomatisch. Hierbei werden zwei Arten des Übertrags unterschieden:

- (*Tabellen*-)Import via DIGMA: Nach einer formalen Prüfung der Feldbezeichnungen werden die zuvor erstellten Tabellen via Skript eingelesen und so ins Zotero übertragen, dass die Datensätze im sog. Importbereich (= *Zotero*-Gruppe) als Einzeleintrag erscheinen.³² Nun sind die in der Tabelle erschlossenen Inhalte ausgefüllt.
 - *Sortierung der Zotero-Einträge*: Es folgt der Übertrag vom Importbereich in die Zielsammlungen, die gemäss Aktenplan strukturiert sind. So können ganze Datencluster kopiert werden.³³ Die Dossierposition im Aktenplan wird in der *Zotero*-Struktur abgebildet, wobei die Sammlung der *Zotero*-Gruppe die jeweiligen Dossiers benennt. Die Provenienz der Datensammlung ist in einem separaten Dokument vermerkt.
 - *Erweiterung der Zotero-Einträge*: Sodann werden die Medien, die Zwischenzeitlich vereinnahmt wurden (s. u.) angehängt, indem die Links als Anhang in den Datensatz eingetragen werden. Eine Liste wird vom DIGMA bereitgestellt.

Metadatenquelle Indexer: Im Falle des *Indexer*-Workflows werden die *Zotero*-Einträge automatisch in dem Moment erzeugt, wo der Datensatz die Inventarnummer erhält. Die jeweiligen Medien/Ressourcen hängen als Link (Anhang) bereits an den Datensätzen und sind nummerierten. Zudem sind die Einzelfiles mit einem Auszug der technischen Metadaten aus der Erkennungskaskade des *Indexers* ausgestattet (vgl. Abbildung 4, Seite 117). Semantischen Informationen wie Autor*in, Titel, Datum etc. müssen hingegen händisch eingetragen werden.

- *Sortierung der Zotero-Einträge*: Das Verschieben der Einträge in Unterordner erfolgt ebenfalls händisch, zumal die Verzeichnisstruktur der *Zotero*-Subfolder durch die erschliessende Person erstellt wird.

Erschliessung

Als Referenz für die Erschliessung dient der *Dublin Core*,³⁴ auch wenn die Erschliessungstiefe und -qualität variiert. Zu den Kernmetadaten gehören neben Informationen zu Sammlungskonvolut und dem jeweiligen Erschliessungscluster Angaben zu Autor*in, Titel, Jahr, wenn möglich Ort sowie weitere Metadaten wie etwa Schlagwörter.

31 <https://www.handle.net/> [Abruf 4. 10. 2019].

32 Die Metadaten verbleiben in einer MySQL-Datenbank, die als Gelenk fungiert.

33 Nachdem alle Datensätze in Zielsammlungen kopiert wurden, werden die Inhalte des Importbereichs gelöscht.

34 DCMI – Dublin Core Metadata Initiative: DCMI Metadata Terms. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> [Abruf 4. 11. 2019].

Zugriffsrechte werden in der Regel abhängig von der jeweiligen Sammlung (Dossier) geregelt.³⁵ So definieren z. B. die Institute als originäre Sammlungseigner den Vertraulichkeitsstatus einer Dokumentengruppe/Dossiers. Die Entscheidung über die Zugänglichkeit wird i. d. R. durch die Institutsleitung oder ihr Sekretariat an die Mediathek kommuniziert, die diese in Absprache mit dem DIGMA umsetzt. Zur Auswahl stehen hierbei die Stufen:

- Freie Nutzung im Internet
- Nutzung innerhalb des Schweizer Hochschulnetzes (Shibboleth-Zugriff)
- Nutzung innerhalb der FHNW
- Nutzung für HGK Angehörige
- Nutzung für Institutsangehörige
- Nutzung für einzelne Jahrgänge eines Instituts plus z. B. Leitung
- Zugriff durch selektierte Einzelpersonen.

Längerfristig besteht das Ziel darin, alle semantischen Beschreibungsdaten frei im Internet zugänglich zu machen, auch wenn die eigentlichen (Meiden-)Daten, also PDFs, Video-, Audio-, Bilddaten nur definierten Nutzer*innengruppen zugänglich sind oder (zeitliche) Embargos gelten. Von einigen Einträgen und Beständen dürfen allerdings auch die Metadaten nur im internen Bereich sichtbar sein.

Medienverarbeitung / Zusammenführung

Nach der Erschliessung spaltet sich die Bearbeitung von Meta- und Mediendaten auf, sodass zwischen dem Import der semantischen Zotero-Metadaten in den SOLR-Index (Katalog-Workflow), den technischen Workflows des Media-Servers (Transkodierungs-Workflows) und den Archivierungs-Workflows unterschieden wird.

Katalog-Workflow

Um die Recherche und eine einheitliche Wiedergabe über die unterschiedlichen Bestände und Sammlungen zu ermöglichen, werden die technischen und semantischen Metadaten (inkl. Handle) im *Integrierten Katalog* auf sogenannten Detailseiten sammlungsspezifisch ausgegeben (vgl. Abbildung 5 und 6, Seite 118–119).³⁶ Hierzu werden die Metadaten zunächst in den SOLR-Volltextindex importiert. Es folgt ein Mapping zwischen den spezifischen Metadaten der neu importierten Sammlung und den bereits existierende Wiedergabefolien, sodass ein wiedererkennbares Erscheinungsbild auch dann entsteht, wenn sehr heterogene Quellen vorliegen.

Zumeist erfolgen nach der ersten Bereitstellung der Resultate im Webinterface weitere Anpassungszyklen, in denen Fehler oder Sonderwünsche auf allen Ebenen behoben respektive implementiert werden. Zudem können Inhalte weiterhin erschlossen oder die Granularität der Sichtungskluster modifiziert werden.³⁷

35 Neu werden die Nutzungsmetadaten (Copyright-Holder, Lizenz, Zugangsform) der Medien zudem direkt ins File geschrieben.

36 Weitere Ansichten finden sich in den Folien des zugehörigen Vortrags der AUDES 2019 online unter: <https://irf.fhnw.ch/handle/11654/27477> [Abruf 4. 10. 2019].

37 Bisher wurden Modifikationen lediglich in der gleichen Erschliessungsperiode eingebracht. Für Veränderungen, die zeitlich oder semantisch versetzt sind und damit als spätere/an-

Transkodierungs-Workflow

Beim Einlesen der Mediendateien in den Mediaserver werden die Dateinamen in die Datenbank des Mediaservers eingetragen und eine Kopie bzw. ein Verweis auf die Mediendatei erzeugt. Zudem werden die technischen Metadaten mit FFmpeg* extrahiert, sodass die Länge eines Videos, das Seitenverhältnis, die Auflösung, der Farbraum, der Container und der Codec etc. bekannt sind und als technische Metadaten in der Datenbank gespeichert werden können. Es folgt die Transkodierung, bei der web- und archivtaugliche Zielformate erzeugt werden:

- *Bilder* werden im Mediaserver in beliebige Grössen in alle gängigen Formate ausgeliefert und intern transkodiert. Sehr grosse Bilder können über einen IIF-Server ausgeliefert (*Cantaloupe*) werden.³⁸
- *Videodaten* werden aktuell in Originalauflösung beibehalten und in ein webtaugliches H.264 Video im MPEG4 Container transkodiert,³⁹ aus dem 25 äquidistante Screenshots im PNG-Format erzeugt werden.
- Von *Audiodaten* werden Sonogramme im SVG-Format erzeugt und ein webtauglicher MPEG4 Container mit AAC Audio (384 kbit, 44100 Hz) wird erzeugt. Ferner wird eine im OPUS-Codec kodierte Audiodatei mit 384 kbit, variabler Bitrate und Kompressionslevel 5 im WebM Container vorgehalten.⁴⁰

Alle Streaming-Dateien werden als Derivat (d. h. mit einem Verweis auf das Eltern-Element der Quelle) vom Mediaserver erneut vereinnahmt, um dann Wiedergabepakete (DIP)⁴¹ je nach Zugriffsart in Echtzeit in den jeweils gewünschten Auflösungen, Grössen und Medienformaten ausliefern zu können.

- PDFs, Audio- und Video-Daten werden (analog zu den grossen Plattformen Youtube, Vimeo etc.) via iFrame mit Player ausgeliefert.
- Alle weiteren Formate werden je nach Projektverlauf abhängig vom Inhalt und der Frage der Zugänglichkeit bearbeitet. Betroffen sind hier häufig raw- oder Produktionsformate (z. B. PSD, INDD, ProRes....) aber auch Spezialformate wie etwa GPX (z. B. beim *Grenzgang*-Projekt), 3D-Daten oder obsoletere Formate.

.....
 dersartige Zutat wie in einer weiteren Metadaten-schale kenntlich gemacht werden sollten, sind wir am Erarbeiten entsprechender Werkzeuge. Denkbar wäre eine Zusammenarbeit mit der UB-Freiburg, welche mit CiTAR bereits Konzepte zur stabilen Referenzierung emulierter Daten erarbeitet hat, vgl. Rechert, Klaus: Preserving Virtual Research Environments – Introducing CiTAR (Part 1) – Open Preservation Foundation, 2018. Online available <https://openpreservation.org/blog/2018/12/12/preserving-virtual-research-environments-introducing-citar-part-1/> [Abruf 4. 10. 2019].

38 Cantaloupe Image Server: <https://cantaloupe-project.github.io/> [Abruf 4. 10. 2019].

39 Videoencoding: Container: MP4 (isom/512), Metadata at start of file (faststart); Video: Encoder: x264; Codec: H264 / Profile: High / Level: 3.1; Preset: Slow; Scene detection: no; Decoder Buffer size: 2200k; GOP Size: 100; Framerate: 25fps; Colorspace: yuv420p; Bitrate: ca. 1100kb/s; Optimierungen: Deblocking (-1:-1), psychovisuelle (1.0:0.15).

40 Audio: Encoder: fdk_aac; Codec: aac; Kanäle: Stereo; Abtastrate: 48000Hz; Bitrate: ca. 196kb/s.

41 Gemäss OAIS, vgl. nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie, und Deutsche Nationalbibliothek. 2013. „Referenzmodell für ein Offenes Archiv-Informationssystem – Deutsche Übersetzung“, 2013, Abschn. Version 2.

Wiedergabe

Je nach Bestand wird eine sammlungsspezifische Startseite generiert, welche die Inhalte thematisch beschreibt. Ferner zeigt die Seite zumeist eine für den Bestand relevante Abbildung, listet sogenannte Einstiegslinks auf spezifische Suchergebnisse oder Detailseiten der Sammlung (z. B. die jeweiligen Künstler*innen) auf und enthält häufig einen Suchschlitz für die gefilterte Suche.

Archivierungs-Workflow

Im Mediaserver können die Medien (Originale) nicht nur transkodiert werden, sondern die Quelldaten können mit den transkodierten Medien sowie den Metadaten zu Archival Information Packages (AIPs) gebaut werden. Hierzu werden die Originaldateien, die Derivate sowie sämtliche technische und semantische Metadaten (als XML-File) in einen BagIT Archivcontainer geschrieben. Es ist vorgesehen, dass die technischen und die semantischen Metadaten ab 2020 in einem METS*-Container dauerhaft aufbewahrt werden.⁴²

Hinzu kommt für diese AIPs ein regelmässiges Reporting. Als Sonderwerkzeug des Archivierungs-Workflows wurde daher *TBBS – The Boring BagIT Store* entwickelt,⁴³ das AIPs oder Gruppen von AIPs an verteilten Orten speichern kann. Aufgrund der potentiell variierenden Sicherheit dieser Speicherorte können die Daten vor dem Transfer verschlüsselt werden. Auch können die Speicherorte bezüglich ihres Sicherheitslevels klassifiziert werden, sodass Inhalte, welche beispielsweise an weniger vertrauenswürdigen Speicherorten aufbewahrt werden, häufiger abgelegt werden. Das ermöglicht eine modulare Kosten-Sicherheits-Politik. Mit Blick auf die Datensicherheit wird zwischen der Sicherheit in der Datenspeicherung und der Sicherheit bezüglich des Datenzugriffs unterschieden. Alle Speicherschritte sowie die Informationen zu den gespeicherten Inhalten werden in umfassenden, automatisch generierten Reports dokumentiert.

Fazit – oder: was bleibt

Niederschwelliges, verteiltes Arbeiten, Nutzerzentrierung und Kosteneffizienz sind grundlegende Aspekte, die den System- und Software-Entscheidungen des digitalen Archivs der HGK zugrunde liegen. Die ausgesprochen schlanke Erschliessungslösung mit *Zotero* wird durch das *DIGMA* und seine Programmierleistungen ebenso möglich, wie die Einrichtung spezifischer Workflows oder die fließende Anpassung an sich wandelnde Kontexte. So können alle Softwarebibliotheken und Systeme fallspezifisch eingerichtet oder ausgebaut werden. Nachhaltig wird das digitale Archiv durch die Speichersysteme, die Schlichtheit und Überprüfbarkeit der Ablage sowie das *TBBS*-System mit

42 Der automatisierte METS-Workflow soll voraussichtlich ab Frühjahr 2020 zur Verfügung stehen. Sofern dies nicht möglich sein sollte, werden die semantischen Metadaten als XML-File exportiert, sodass eine Datenbank-unabhängige Zusammenführung von technischen, semantischen Metadaten und Medien auch im Havariefall oder im Falle eines Systemtransfers möglich sein wird.

43 Enge, Jürgen; Lurk, Tabea: *TBBS – THE BORING BAGIT STORE: Storing digital art and cultural assets in a distributed storage infrastructure*, in Bearbeitung.

seinen Integritätsprüfungen, die explizit für den Archivierungszusammenhang entwickelt wurden.

Die Kulturwissenschaftlerin Simone Osthoff hat vor nicht allzu langer Zeit mit Blick auf aktuelle Veränderungen in Archiven der Gegenwartskunst beobachtet: «when the past refuses to settle and technology designed to upgrade and perpetuate itself further keeps the archive alive; it seemed to me that if not alive, the archive in contemporary art is in a state of flux, undead, ghostly».⁴⁴ So bedrohlich das *Im-Fluss-Bleiben* im klassischen Archivkontext erst einmal klingen mag, weil ja Stabilität und Dauerhaftigkeit installiert werden sollen, scheint hier doch ein wahrer Kern enthalten zu sein, der sich im Hinblick auf unsere Arbeit durchaus konstruktiv lesen lässt: So versuchen wir tatsächlich, dem *Im-Fluss-Bleiben* (flux) der digitalen Daten und Entwicklungen mit einer flexiblen Archivstruktur und einem für Inspirationen offenen Denken zu begegnen. Unser Ziel ist es dabei, Nachhaltigkeit zu schaffen ohne die vorhandenen und für Gestaltung und Kunst so wichtigen und einzigartigen Dynamik allzu massiv einzugrenzen. Untot (undead) oder lebendig zu halten, bleibt gerade mit Blick z. B. auf die Formatvielfalt eine stetige Herausforderung, die wir im vorliegenden Text ebenso ausgeklammert haben, wie das Vermeiden von verwaisten Daten, die quasi *geisterhaft* (ghostly) und quellenlos durchs Archiv spuken und in fast jedem, zumindest in natürlich gewachsenen, digitalen Beständen vorhanden sind. Beide Themen sind uns bewusst, was sich unter anderem im Kernmetadatenatz von Neuzugängen zeigt, der nur noch Rechte-, Lizenz, ggf. Embargo und Copyright-Angaben erhalten muss.

Das Preservation Management wird im vorgestellten System vornehmlich durch den Mediaserver ausgeführt, der Dank *TBBS* dokumentierte Speicher- und Kontrollroutinen ermöglicht. Derzeit wird zudem die gesamte Systemarchitektur inkl. *Indexer*-Routinen auf Micoservices umgestellt, wodurch ein zukunftsfähiges, einfach skalierbares und in den technologischen Teilkomponenten wandelbares System entsteht.

Dass Sammlungsbestände auch oder gerade nach der Archivierung potentiell im wissenschaftlichen Kontext nachgenutzt, erweitert oder (in Derivaten) modifiziert werden können, zeigt die Nähe zu Forschungsdatenzyklen, deren Natur im kontinuierlichen Weiterschreiben und in der Ermöglichung des Informationskontinuums liegt.⁴⁵ Hier greift ferner nach wie vor das von Andrew Treloar et al. entworfene Konzept der *Curation Boudaries*, das den Statuswechsel von Dokumenten beim Schwellenübertritt (Publikation) beschreibt, also in dem Moment, wenn z. B. neue Nutzer*innengruppen Zugriff erhalten.⁴⁶ Dabei kommt es zu einer Festschreibung der bestehenden Metadaten, sodass Änderungen und Erweiterungen möglich werden, ohne dass der Quellenstatus gefährdet wird.⁴⁷

44 Ludovico, Alessandro: Interview [with] Simone Osthoff, in: Neural No. 58 (2017: hacktivism), S. 7. Das Gespräch bezieht sich auf Ostoffs ebenfalls 2017 erschienenes Buch *Performing the Archive. The Transformation of the Archive in Contemporary Art* der Interviewten.

45 Digital Curation Centre: DCC Curation Lifecycle Model. 2019. <http://www.dcc.ac.uk/resources/curation-lifecycle-model> [Abruf 4. 10. 2019].

46 Treloar, Andrew, David Groenewegen, und Cathrine Harboe-Ree: *The Data Curation Continuum: Managing Data Objects in Institutional Repositories*. D-Lib Magazine 2007. 13 (9/10). <https://doi.org/10.1045/september2007-treloar> [Abruf 4. 10. 2019].

47 Vgl. Treloar, Andrew: *Presentations: Diagrams*. 2017. Online available <https://andrew.treloar.net/research/diagrams/> [Abruf 4. 10. 2019].

Das Archiv der HGK befindet sich daher in einer immerwährenden Weiterentwicklung. Es ist kritisch in seiner Selbstreflexion und beabsichtigt, sich durch den Austausch mit der Fachcommunity stets zu verbessern. Die Grundhaltung orientiert sich ein Stück weit an Konzepten des *verteilten Wissens* oder des *Networks of Care*, wie es im Kontext der Konservierung, Restaurierung und Archivierung von digitalen Kunstwerken seit geraumer Zeit diskutiert wird.⁴⁸

Summary

The article presents the most important indexing workflows and tools of the digital archive of the Academy of Art and Design HGK FHNW Basel. Since the data are very heterogeneous in terms of media technology, structure, source, and quality (in depth of indexing), the collection and indexing processes adapt to the respective collections and contexts rather than vice versa. The renunciation of strict archive specifications or a singular filter requires a broad range of different tools, because despite the preservation of semantic, technical or cultural-aesthetic conventions and preferences, search and access should later be able to take place systematically and comprehensibly.

The text begins with a brief overview of the archival collections, which makes the reciprocal requirements of the data suppliers and the archives transparent. The process design and the sequence of the indexing routines can be derived from this relationship. Since all workflows are linked to digital tools, these are then presented. The outlook reflects on aspects that might be transferred to other respective archives.

48 Vgl. hierzu Dekker, Annet: Enabling the Future, or How to Survive FOREVER. In: A Companion to Digital Art (Wiley Blackwell Companions to Art History, 9), hrsg. von Christiane Paul, Chichester: Wiley Blackwell, 2016, S. 553–574.

Abbildungen

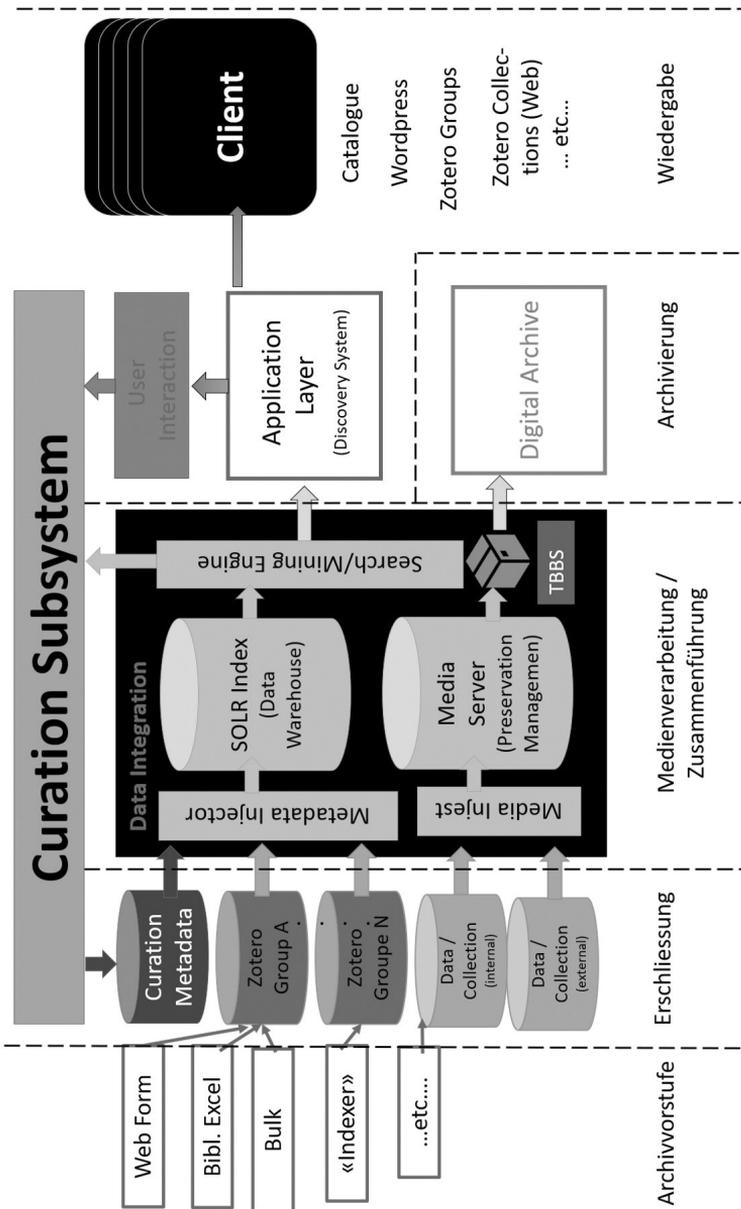


Abb. 1: Schematische Darstellung der Systemarchitektur. Zu unterst sind die unterschiedlichen Archivierungsschritte im zeitlichen Verlauf modellhaft gekennzeichnet.

The screenshot shows the Zotero interface. On the left, a list of documents is displayed, including titles like 'Schriftliche Arbeit', 'Bünzli Adrian', 'Bürgi Yolanda', 'Disler Shirin', 'Egger Simon', 'Hunter Chris', 'Lengsfeld Lena', 'Meyer Alexandra Maril', and 'Müller Rebekka'. The right pane shows the metadata for a selected entry, including fields like 'Eintragsart', 'Titel', 'Autor', 'Mitarbeiter', 'Zusammenfassung', 'Art', 'Ort', 'Datum', 'Anzahl der Seiten', 'Sprache', 'Kurztitel', 'URL', 'Heruntergeladen am', 'Archiv', 'Standort im Archiv', 'Bibliothekskatalog', 'Signatur', 'Rechte', 'Extra', 'Hinzugefügt am', and 'Geändert am'.

Titel	Ersteller	hinz...
Wozu Kunst?	Bünzli	18.7.201...
2 x; 1 x EM00...		24.7.201...
Bünzli Adrian...		19.7.201...
Kitsch?	Bürgi	18.7.201...
Zwischen Bild u...	Disler	18.7.201...
Dada - ein Selb...	Egger	18.7.201...
Vom sammeln u...	Hunter	18.7.201...
Scham	Lengsfeld	18.7.201...
Intimität in der ...	Meyer	18.7.201...
Animation. Mal...	Müller	18.7.201...
Die Frau als Mus...	Ott	18.7.201...
Mise en scène d...	Pitteloud	18.7.201...
Louise Bourgeoi...	Savic	18.7.201...
Schmerzensmann	Stucky	18.7.201...
Ein Überblick üb...	Studerus	18.7.201...
Cindy Sherman	Vukovic	18.7.201...
Treppen in der ...	Willa	18.7.201...

Abb. 2: Zotro-Workflow: Ganz links eine Fileliste als Word-Liste; die Inhalte werden in ein Excel eingetragen (fehlt in Abb.). Mittig und rechts ist dann der Zotero-Eintrag zu sehen. Die Inventarnummer hier (Feld: Extra) entspricht dem Bibliotheks-Strichcode auf der Archivbox dieser spezifischen Sammlung.

Home
Clone Window

Intelligent Read-Only Media Identification Engine

Iron Maiden - The Indexer

11 Ergebnisse (0 Sekunden) [file inventory "B0000077928"]

Prev
1
2
Next
Page

	act	act Basel	#4.4419.2213096	2018-08-20T16:45:53Z		#B0000077928	Exchangeable Image File Format (Compressed)
	JPG	DSC_1417.JPG	image/jpeg	672.09 KiB	2018-08-20T08:44:38Z	fmi645	Exchangeable Image File Format (Compressed)
/act/act07/Fotos/2007_Foto_act_Base/Osqual/Sarina Scheidegger, Anina Krieger_HGK_Basel							

	act	act Basel	#4.4419.2213096	2018-08-20T16:45:53Z		#B0000077928	Exchangeable Image File Format (Compressed)
	jpg	DSC02961.jpg	image/jpeg	13.09 KiB	2018-08-20T08:44:38Z	x-fmt391	Exchangeable Image File Format (Compressed)
/act/act07/Fotos/2007_Foto_act_Base/Osqual/Sarina Scheidegger, Anina Krieger_HGK_Basel							

	act	act Basel	#4.4419.2101204	2018-05-05T17:57:11Z		#B0000077928	Exchangeable Image File Format (Compressed)
	JPG	DSC_1421.JPG	image/jpeg	804.65 KiB	2018-08-20T08:44:38Z	fmi645	Exchangeable Image File Format (Compressed)
/act/act07/Fotos/2007_Foto_act_Base/Osqual							

	act	act Basel	#4.4419.2213087	2018-08-20T16:45:53Z		#B0000077928	Exchangeable Image File Format (Compressed)
	JPG	DSC_1421.JPG	image/jpeg	804.65 KiB	2018-08-20T08:44:38Z	fmi645	Exchangeable Image File Format (Compressed)
/act/act07/Fotos/2007_Foto_act_Base/Osqual/Sarina Scheidegger, Anina Krieger_HGK_Basel							

content
file
libmagic
gvis-info
tika
imagemagick
siegfried
cite this item

mimetype: image/jpeg **minencoding:** binary **description:** JPEG image data, Exif standard: [TIFF image data, big-endian, direntries=13, descriptio= solutionunit=2, software=QuickTime 7.1.3, datetime=2007:05:24 14:36:46, hotcomputer=Mac OS X 10.4.9], comment: "AppleMark", baseline, precision 8, 32bx 246, frames 3

deutsches
literatur
archiv marbach

University of Applied Sciences and Arts Northwestern Switzerland
Academy of Art and Design

Abb. 3: Indexer-Resultate des ACT-Bestands, nachdem die Inventarnummer vergeben wurde. Detail (zweiter Eintrag von oben): libmagic-Analyse.

Zotero

Titel

- la grande bouffe
- There must be a corner
- mi le tempo
- Beat Boxer
- Porter Basel
- Son Goku and the Super Sky...
- Homotransmitter
- Handschin
- Lustenberger
- Merz
- Murmann
- Roby
- Die Unmöglichkeit des Fliehe...
- sare titre
- Fritou
- Le fi:deux pièces
- http://act-perform.net/200...
- 001/010 [44.4419.2213087...
- 002/010 [44.4419.2213088...
- 003/010 [44.4419.2213089...
- 004/010 [44.4419.2213090...
- 005/010 [44.4419.2213091...
- 007/010 [44.4419.2213093...
- 008/010 [44.4419.2213094...
- 009/010 [44.4419.2213095...
- 010/010 [44.4419.2213096...
- ACT Website
- Toast cets
- wisser Wasserfischenthal
- Programm
- Publium/Event
- Trailer
- Videovorschau DVD-Menu
- Vorbereitung Interviews

Ersteller

- Baino
- Baumgartner
- Billari
- Chiquet und Re...
- Chiquet et al.
- Grossenbacher ...
- Handschin
- Lustenberger
- Merz
- Murmann
- Roby
- Roquero
- Routnova
- Ruffieux
- Schledegger un...
- http://act-perform.net/200...
- 001/010 [44.4419.2213087...
- 002/010 [44.4419.2213088...
- 003/010 [44.4419.2213089...
- 004/010 [44.4419.2213090...
- 005/010 [44.4419.2213091...
- 007/010 [44.4419.2213093...
- 008/010 [44.4419.2213094...
- 009/010 [44.4419.2213095...
- 010/010 [44.4419.2213096...
- ACT Website
- Toast cets
- wisser Wasserfischenthal
- Programm
- Publium/Event
- Trailer
- Videovorschau DVD-Menu
- Vorbereitung Interviews

Titel der Reihe

- Le fi:deux pièces

Band

- # von Bänden
- Ort
- Ortstag Basel

Datum

- 2007-05-06

Sprache

- Mediathek HGK

ISBN

- Mediathek HGK

Kurztitel

- Archiv

Standort im Archiv

- Act

Bibliothekskatalog

- Signatur B000007928
- URL http://hdl.handle.net/201500.11806/mediathek/inventory/6000007928

Heruntergeladen am

- Rechte
- Extra
- Hinzugefügt am 23.2.2018, 15:28:17
- Geändert am 8.4.2019, 14:17:53

Verwendete Einträge: [hier klicken]

Tags: [hier klicken]

Absatz:

Bestand 4: Act
 Session 4419: act Basel
 File 2213087: DSC_1421.JPG
 Path: /act/act07/Fotos/2007_Foto_act_Basel_Ostquai/Sarina_Schledegger_Arma
 Kriegler_HGK_Basel

Image:
 Mimetype: image/jpeg
 Encoding: binary
 Description: JPEG image data, EXIF standard: [TIFF image data, big-endian, dntres=10, manufacturer=NIKON CORPORATION, model=NIKON D70, orientation=upper-left, x-resolution=168, y-resolution=176, resolutionunit=2, software=Ver.1.02., dateime=2007:05:05 20:34:27], basette, precision 8, 2540x1488, frames 3

GFJS:Info:
 Mimetype: image/jpeg

Seggfried:
 Pronom-ID: fmf/645
 Format: Exchangeable Image File Format (Compressed)
 Version: 2.2.1
 Mimetype: image/jpeg

Image Mapack:
 Format: JPEG; Geometry: 2540x1488; xres: 300; yres: 300;

Apache Tika:
 Mimetype: image/jpeg
 [Exposure Sequence Number]: 3675
 [EXIF Version]: 2.21
 [Unknown Z]: 0
 [Number of Pages]: 1
 [Sensor Pixel Size]: 7840 7840
 [Number of Components]: 3
 [CFA Pattern]: [Blue,Green][Green,Red]

Abb. 4: Zotero-Eintrag eines Indexer-Objektes: Links die Sammlung, Mitte die semantischen Metadaten, Rechts die Indexer-Daten des Files 001/009 der Inventarnummer B0000078035

Technische Universität
Hochschule für Gestaltung und Kunst

Institut Kunst

Mediathek

Aktuelle Objekte

Binzli, Adrian ; Kurzweyer, Roman
Wozu Kunst? 2011
Basel

Note

2 x 1 x EM000008482411 (X_014_1)

Binzli, Adrian, TA.pdf

Thesen

BA 1
Schriftliche Arbeit Basis Thesis
Institut Kunst
Textarbeiten BA
2011

Weitere Inhalte aus "Institut Kunst - Textarbeiten BA.2011:BA 1"

Studeris, Silvia et al.	Ein Überblick über die Entwicklung der Kunstgeschichte
Stucky, Raphael et al.	Schmerzemann
Savyc, Jelena et al.	Louise Bourgeois. Welche Rolle spielt der Körper in der Kunst?
Bürgi, Johanda Esther et al.	Kinship?
Vukovic, Meri et al.	Cindy Sherman
Erger, Simon et al.	Dada - ein Selbstversuch
Müller, Rebekka et al.	Annäherung. Malerei in Bewegung. Untersuchung anhand von zwei Arbeiten von William Kenridge und Caroline Leaf
Hunter, Chris et al.	Vom sammeln und (er)finden. Ein Reisebericht bis Christian Boltanski. 05.07.2011 - 16.08.2011
Ott, Bianca et al.	Die Frau als Muse und Schöpferin. Oder: Wie Frauen dargestellt werden - eine Untersuchung über nackte Inszenierung des weiblichen Körpers in der Kunst
Düster, Shirin et al.	Zwischen Bild und Raum. Zur Transformation des Rahmens und der Rolle des Betrachters
Pinefold, Moira	Muse en scène de l'art, m'écrit: quelques réflexions sur la visite de l'exposition "Eloge du doute"
Willa, Tobias et al.	Treppen in der Kunst
Lengsfeld, Lena et al.	Scham

Mezzar, Alexandra et al. [Intermitt in der Öffentlichkeit](#)

Abb. 5: Detailseite einer Theoriearbeit des Instituts Kunst; korreliert mit Abbildung 2.

ACT-Basel

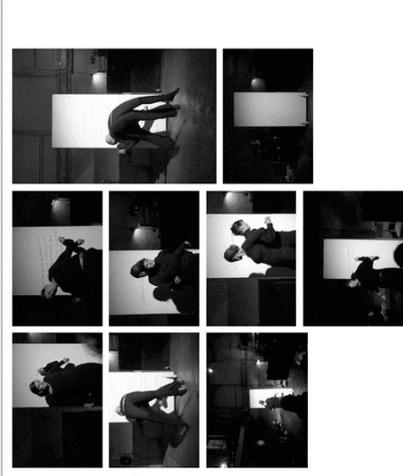
Aktuelles Objekt

Schädinger, Sarina; Krüger, Amina
 Le fi, deux pièces 2007
 Original: Basel

Neu

<http://act-perform.net/2007/basel/sarina-schadinger-amina-krueger>

Gallery



Themen

- Blancieren
- Berühren
- Schreiben
- Hochschule für Gestaltung und Kunst FHAW, Basel
- Abhängigkeit
- Balance
- Erfahrung
- Gegenseitigkeit
- Gleichgewicht
- Sprache
- Unabhängigkeit
- Körper
- Stift
- Wandelment
- ACT-Basel
- 2007
- Basel

ACT Website

<http://act-perform.net/2007/basel/sarina-schadinger-amina-krueger>

Abstract

Wir haben keinen eigenen Hintergrund mehr. Dort wo mein Rücken wäre, bist du. Dort wo deiner wäre, ich. Unabhängig, abhängig, individuell und doch zusammen. Was wäre ein Teil ohne den Anderen? Was wäre ich ohne Sie? Die Stärke des Einen ist die Balance des Anderen. Gewicht unterstützt Gleichgewicht. Ineinander verflochten trägt der Eine den Anderen und der Andere den Einen. Ein Zusammenspiel aus Wille und Leidenschaft, Zweifel und Freude. Wir machen uns eins obwohl es schmerzt. Wir haben Hintergrund.

Verweise

- Z2006003.CIV.ZZEKP
- Z2006003.T52/CBSS

Weitere Inhalte aus "ACT-Basel - 2007:Basel"

Rouznova, Anna	📄	sans titre
	📄	Trailer
Baumgartner, Nino	📄	There must be a corner
Barrin, Aude	📄	la grande bouffe
Chiquet, Fabian et al.	📄	Best Boser
Grossenbacher, Denise et al.	📄	Son Goku and the Super Saig'in
Merz, Jon	📄	Straussenspolitik
	📄	Programm

Abb. 6: Detailseite eines ACT-Eintrags als Resultat des Indexer-Zotero-Workflows; korreliert mit Abbildung 3 und 4.

DIE INDEXIERUNG VON AUDIOVISUELLEN ARCHIVEN UNTER BENUTZUNG VON METHODEN DER AUTOMATISCHEN SPRACH- UND BILDERKENNUNG

Jan Švec, Pavel Ircing

Einleitung

Dank der Erreichbarkeit von kostengünstiger Aufzeichnungstechnik und des beispielslosen Anwachsens der Speicherkapazitäten von Festplatten oder (insbesondere in letzter Zeit) von online-Datenspeicherplätzen ist es gegenwärtig kein Problem, eine beinahe beliebige Menge von audiovisuellen Daten hochzuladen und zu speichern. Die in solchen Daten abgespeicherten Informationen sind jedoch de facto unzugänglich, wenn wir über kein Werkzeug zu ihrer effektiven Auffindung verfügen, denn es ist nicht möglich, die bestehenden riesigen (und sich ständig vergrößernden) Datenvolumen in Sequenzen durchzugehen – und zwar nicht einmal, wenn wir uns auf Aufzeichnungen aus einer wie auch immer eng begrenzten Domäne konzentrieren würden.

Die Funktion „Auffinden von Informationen“ (engl. *information retrieval* – IR) stellte noch vor 30 Jahren eine akademische Disziplin dar, die im Prinzip bloß Bibliothekaren und weiteren mit Informationen arbeitenden Fachleuten wie Journalisten und Juristen vorbehalten war. Mit der Entwicklung des World Wide Web in den 90er Jahren des 20. Jhdts. verschob sich diese Disziplin in das Internet und in den letzten mehr als zehn Jahren ist die Informationssuche im Web praktisch für alle Nutzer von PCs und an das Netz angeschlossenen Mobilgeräten eine Routineangelegenheit. Weiterhin handelt es sich jedoch vor allem um Textsuche. Obwohl einige Suchmaschinen auch die Möglichkeit der Suche in multimedialen Daten anbieten, handelt es sich entweder um eine Suche, die auf der manuellen Beschreibung oder, im Fall der Sprachaufnahmen, auch auf der Transkription der einzelnen audiovisuellen Dokumente begründet ist.

Die automatische Informationssuche in Sprachdaten (d. i. ohne die Notwendigkeit einer händischen Annotation der Aufnahmen) begann Anfang der 90er Jahre des letzten Jahrhunderts in das Zentrum der Aufmerksamkeit der Wissenschaftler zu rücken. Etwa ein Jahrzehnt später erschienen die ersten praktischen Systeme, die in der Lage sind, in konkreten umfangreichen

Audioarchiven zu suchen¹. In den letzten Jahren können wir auch auf einige Suchmaschinen treffen, die fähig sind, mit dem üblichen (und sich andauernd vergrößernden) Bereich von im Internet aufzufindenden Audio- und Videoaufnahmen zu arbeiten.²

Die überwiegende Mehrheit der Suchsysteme in Sprachdaten ist als Serienkombination eines Sprachdekoders mit einem Informationssuchsystem in Textdaten konzipiert. Diese Konfiguration ermöglicht es relativ leicht die schon existierenden Systeme für automatische Spracherkennung (engl. *automatic speech recognition* – ASR) und für Informationssuche (IR) zu benutzen. Auf der anderen Seite wird bei solch einer Kombination von zwei im Prinzip unabhängigen Modulen die Tatsache ignoriert, dass die durchsuchten Textdokumente mit Erkennungsfehlern behaftet sind, aber dass auch ein möglicher „reicherer“ Output des Sprachdekoders zb. in Form von Gittern nicht benutzt wird. Es ist also das Ziel dieses Beitrages, vor allem Methoden vorzustellen, die auf einer engeren Verknüpfung der ASR- und IR-Module beruhen.

Die Indexierung und die Suche im USC Shoah Foundation Visual History Archive

Dieses Archiv entstand auf Anregung des US-amerikanischen Regisseurs Steven Spielberg, der nach der Aufführung seines Filmes Schindlers Liste (1993) die *Survivor's of the Shoah Visual History Foundation* (VHF) gründete. Die Stiftung heuerte Kameramänner und Reporter an, die im Lauf der Jahre 1994–1999 ungefähr 52000 Gespräche in 32 Sprachen aufzeichneten (es handelt sich insgesamt um mehr als 160 000 Stunden Aufnahmen, in etwa die Hälfte der Gespräche wurde in Englisch geführt). 2006 übersiedelte das VHF (und zwar sowohl physisch, als auch administrativ) an die University of Southern California (USC) und wirkt seither unter der Bezeichnung *USC Shoah Foundation – The Institute for Visual History and Education*³

Es war die ursprüngliche Absicht, das ganze Archiv manuell zu bearbeiten – die Aufnahmen in Segmente aufzuteilen, die immer einem Thema gewidmet sind und jedes solches Segment mit Folgendem zu versehen:

- einer Kurzzusammenfassung – drei das Segment beschreibende Sätze
- einem aus dem Thesaurus ausgewählten Grundthema, das cca. 3 000 Konzepte/Themen enthält – etwa Familienleben, Bildung, Transporte in die Konzentrationslager u.Ä.
- einer geographischen Verortung entnommen aus einem Verzeichnis aus 30 000 Ort-Jahreszahl-Paaren, etwa Berlin-1939, Praha-1942
- Namen der Personen, über die im entsprechenden Segment gesprochen wird

1 Hansen, John H. L. et al.: SpeechFind: Advances in Spoken Document Retrieval for a National Gallery of the Spoken Word. In: IEEE Transactions on Speech and Audio Processing 13 (2005), 712–730

2 Goto, Matasaka; Ogata, Jun: PodCastle: Recent advances of a spoken document retrieval service improved by anonymous user contributions. In: Proceedings of Interspeech 2011 (2011), 3073–3076

3 <https://sfi.usc.edu/> [Abruf 4. 10. 2019]

Auf diese Weise wurden 4 000 auf englisch geführte Gespräche bearbeitet (cca 10 000 Stunden Aufnahme, 8% des Archivs). Im Hinblick darauf, dass diese Annotation 150 000 Arbeitsstunden erforderte und 8 Mio. USD kostete, kamen die Archivbetreuer zum Schluss, dass es zeitlich und finanziell zu anspruchsvoll wäre, das ganze Archiv auf diese Weise zu bearbeiten. Das gab den Anstoß zum Entwurf des Projekts MALACH. Dieses von der US-amerikanischen *National Science Foundation* (NSF) finanzierte Projekt setzte sich zum Ziel, die Methoden der automatischen Spracherkennung und Informationssuche einzusetzen, um den Zugang in das VHF-Archiv mittels einer Beschleunigung und Kostensenkung des Katalogisierungsprozesses zu verbessern. Es war also der ursprüngliche Gedanke, dass man die oben angeführte Katalogisierung – die Zuordnung von Metadaten – auch weiterhin durchführt, aber bloß die händische Arbeit durch maschinelle Methoden ersetzt wird. Im Verlauf des Projekts kam es jedoch zu einer bedeutenden Verlagerung der gesamten Zugangsweise zu der Aufgabe, und zwar vor allem weil die unvollkommene automatische Zuordnung der Metadaten eine große Fehlermenge in den Suchergebnissen verursachte.

Suche nach komplex spezifizierten Anfragen

In einer weiteren Phase wurde entschieden, dass der „Informationsbedarf“ der Nutzer zum Zweck der Suche durch ein komplex spezifiziertes Thema zum Ausdruck gebracht werden soll, dessen Beispiel wir zur Illustration anführen:

```
<top>
<num>3014</num>
<title>Sionismus</title>
<desc>Gespräche über zionistische Bewegungen in Osteuropa, insbesondere über Hashomer-Hatzair-.</desc>
<narr>Gegenstand unseres Interesses sind Geschichten über zionistische Organisationen und Ansichten von Einzelpersonen über den Zionismus sowie die Entscheidungen über einen Eintritt oder Nichteintritt in diese Organisationen. Es werden eigene Erfahrungen bevorzugt, wir haben jedoch auch Interesse an Informationen aus zweiter Hand über die Haltungen und Beteiligung von Verwandten und Freunden in der zionistischen Bewegung.</narr>
</top>
```

Aus einem so spezifizierten Thema wurde mittels einiger grundlegender, bei der Suche in Textdaten benutzter Vorgangsweisen (Lemmatisierung, Entfernung von Stop-Wörtern u.Ä.) eine Frage geschaffen, die in der Folge bei der Suche in mittels der automatischen Spracherkennung gewonnenen Gesprächsprotokollen verwendet wurde. Es war jedoch ein relativ grundlegendes Problem, dass die „klassische“ Textsuche von Informationen mit dem Konzept eines Dokuments arbeitet, d. h. im Idealfall eines thematisch kohärenten Textblocks, den man dem Nutzer gleichsam als Suchergebnis vorlegen kann. In unserem Fall war jedoch das Ergebnis des Sprachdekoders ein zusammenhängender „Textstrom“, dessen einzige Unterteilung aus der Verwendung von 30 Minuten-Videokassetten für die Aufzeichnung der Gespräche hervorgeht. Solch eine Sequenz ist jedoch für eine effektive Informationssuche selbstverständlich zu groß und so

wurde entschieden, künstliche „Dokumente“ in der Weise zu schaffen, sodass ein dreiminütiges Zeitfenster (mit einer einminütigen Überlappung) schrittweise entlang des zusammenhängenden, durch die automatische Sprachtranskription entstandenen Textstromes verschoben wurde und man den in das Fenster „fallenden“ Text immer in ein eigenständiges Dokument übertrug.

Die Suche in solchen „Dokumenten“ war jedoch ziemlich unzufriedenstellend⁴. Ferner bot das erwähnte Suchsystem dem Nutzer nur ein klassifiziertes Identifikatorenverzeichnis der zugehörigen, künstlich geschaffenen Dokumente, deren Lesung – im Hinblick auf die Fehleranfälligkeit des dekodierten Textes von etwa 30% – vom Nutzerstandpunkt nicht akzeptabel ist.

Die oben erwähnten Beobachtungen haben uns auf den Gedanken gebracht, dass strukturierte Anfragen (geschaffen durch Bibliothekare oder Archivare auf der Grundlage der Nutzerbedürfnisse) eher für eine Suche in Daten geeignet sind, die händisch durch Metadaten ergänzt werden (deren Struktur wird letzten Endes auch oft durch dieselben Informatikfachleute entworfen). Insofern wir also die Datenverarbeitung für die Suche so weit als möglich automatisieren und gleichfalls das Archiv direkt dem Endverbraucher zugänglich machen möchten, ist es möglicherweise vorteilhaft, die Zugangsweise zu ändern und dem Nutzer eine Schnittstelle anzubieten, die es ihm ermöglicht, „bloß“ nach dem Vorkommen einzelner Wörter oder kurzer Phrasen zu suchen. Die Intuition sagt uns, dass die Effektivität der Suche (vom Gesichtspunkt der Befriedigung der Informationsbedürfnisse betrachtet) in diesem Fall wenigstens wie bei der Verwendung von komplexen Themen für die Suche vergleichbar sein wird – sei es nur aus dem Grund, dass die Spezifikation der strukturierten Anfragen für den Nutzer unpraktisch ist, oder weil sie mehrheitlich durch die langjährige Verwendung der Internet-Suchmaschinen ohnehin an die Eingabe von Kurzanfragen gewöhnt sind.

On-line-Suche einzelner Wörter und Phrasen

Die in dieser Sektion vorgestellte Methode wurde an unserer Arbeitsstätte für die online-Suche im USC Shoah Foundation Visual History Archive entwickelt⁵ und basiert eher auf heuristischen Vorgangsweisen und einer guten Kenntnis der Datenbank-Problematik als auf den in der IT-Community eingeführten Arbeitstechniken. Dennoch sind ihre Suchergebnisse gut, sie ist sehr schnell und stellt nur geringe Ansprüche an die Speicherkapazität.

Die bei der Datenverarbeitung eingeführte Methode funktioniert in zwei unterschiedlichen Modi – der Indexierung und der eigentlichen Suche.

.....
4 Pecina, Pavel et al: Overview of the CLEF-2007 Cross-Language Speech Retrieval Track. In: Lecture Notes in Computer Science 5152 (2008), 674-686.

5 Psutka, Josef et al: System for fast lexical and phonetic spoken term detection in a Czech cultural heritage archive. EURASIP Journal on Audio, Speech, and Music Processing 2011 (2011),1-10

Indexierung

Für eine effektive Suche (und zwar sowohl vom Gesichtspunkt des Speicherbedarfs, als auch der Geschwindigkeit) ist es naturgemäß entscheidend, Daten, die wir durchsuchen möchten, in einer geeigneten, Index genannten Struktur zu repräsentieren. Wie jedoch schon im vorhergehenden Text angedeutet wurde, möchten wir im Hinblick auf die relativ hohe Fehlerquote der automatisch generierten Texttranskription mit einer mehr komplexen Repräsentation des Sprachdeko-Outputs, dem Gitter, arbeiten. Das Gitter ist formal genommen ein gerichteter azyklische Graph, dessen Kanten Wörter, Gewichtungen, etwa den Zuverlässigkeitsgrad, und eventuell auch weitere relevante Informationen enthalten.

Unsere Methode indexiert das Wortgitter in folgender Weise:

- sie ignoriert alle Kanten, deren Zuverlässigkeitsgrad (durch A-posteriori-Wortwahrscheinlichkeiten ausgedrückt) niedriger als die gewählte Schwelle $\Theta_w = 0.05$ ist.
- sie speichert die übrigen Kanten in der Datenbank (im Index) in Form der Fünfergruppe $(t_s, t_e, \text{word}, p_{\text{word}}, \text{video_id})$ bei der t_s und t_e die Anfangs- und Endzeit des Wortes word ist, das zur gegebenen Kante gehört, und p_{word} für die erwähnte A-posteriori-Wahrscheinlichkeit steht. Der Indikator video_id bezeichnet dann die Datei mit der Videoaufnahme (t_s und t_e sind relativ auf den Anfang dieser Datei bezogen).
- sie durchtrennt den Index so, dass aus der Potenz der mit demselben Wort word bezeichneten Kanten, die sich entweder überdecken, oder voneinander weniger als $\Delta t_w = 0.5$ Sekunden entfernt sind, nur die Kante mit dem höchsten Wert p_{word} beibehalten wird.

Auf diese Weise verlieren wir zwar die Information über die strukturellen Eigenschaften des ursprünglichen Gitters, auf der anderen Seite gewinnen wir jedoch für die Suche einen kompakten und effektiven Index, der dennoch eine viel reichere Informationsbreite enthält, als die Texttranskription, die in den Experimenten verwendet wurden, die in Kapitel 2.1 beschrieben sind.

Suche

Der Suche dient eine Webapplikation, die den Nutzern eine komfortable Schnittstelle anbietet. Solch eine Lösung besitzt eine ganze Reihe von Vorteilen:

- eine beliebige Erreichbarkeit, wo auch immer sich Internet befindet
- dem Nutzer genügt ein geläufiger moderner Internetexplorer
- das Gerät, auf dem der Nutzer arbeiten wird, kann sowohl ein PC, als auch zb. ein Tablet sein
- es ist nicht notwendig, irgendetwas auf dem Nutzergerät zu installieren (außer dem eigentlichen Internetexplorer, der jedoch schon auf beinahe allen Geräten vorinstalliert ist)
- Plattformunabhängigkeit (die Applikation funktioniert unabhängig vom Betriebssystem des Geräts)

- allen Nutzern steht die gleiche Version der Applikation zur Verfügung
- die Aktualisierung der Applikation verläuft am Server (sobald also die Applikation aktualisiert wird, haben alle Nutzer diese aktualisierte Version zur Verfügung)

Das Aussehen der Webseiten mit den Ergebnissen ist in Abbildung 1 dargestellt. Die Seite enthält vier für die Steuerung der Applikation wichtige Elemente. Im oberen Teil ist das insbesondere das Eingabefeld für den Suchbegriff, im linken Teil das geordnete Verzeichnis mit den Ergebnissen der Suchanfrage. Den größten Teil nimmt das Detailergebnis mit dem Videofenster und der Auflistung der zugänglichen Metadaten ein.

Abb. 1, Seite 127

Das Eingabefeld dient der textlichen Eintragung des Suchbegriffs. Ebenso wie die Suchmaschinen von Google oder Seznam unterstützt es spezielle Operatoren, die eine genauere Spezifizierung der gesuchten Wörter ermöglichen. Hauptzweck dieses Feldes ist die Möglichkeit, eine Liste der gesuchten Wörter einzugeben. Diese werden nach der Einleitung der Suche automatisch dekliniert. In einigen Situationen ist die automatische Deklinierung unerwünscht, und daher ist es möglich, das gesuchte Wort mit Hilfe von Anführungszeichen zu „ummanteln“. Alle Wörter (dekliniert oder undekliniert) sind sog. Pflichtbegriffe. Das bedeutet, dass alle in einem bestimmten Zeitintervall gefunden werden müssen, damit das Ergebnis als gültig anerkannt wird. Im Bereich der erweiterten Suche ist es möglich, dieses Intervall einzustellen. Mittels des +-Zeichens am Anfang des Wortes kann man dieses als Nicht-Pflichtbegriff kennzeichnen.

Die gefundenen Ergebnisse werden in einer Trefferliste im linken Teil der Applikation dargestellt. Die einzelnen Posten stellen den gefundenen Text, die Bezeichnung des Ursprungsvideos und den Zuverlässigkeitsgrad dar. Man kann die abgebildeten Ergebnisse nach dem Zuverlässigkeitsgrad, nach dem Datum oder der Zeit ordnen. Nach dem Klick auf einen beliebigen Posten in der Trefferliste wird das entsprechende Detailergebnis abgebildet.

Die Hauptfunktion des Videofensters ist das Abspielen des ausgewählten Abschnitts (der dem gefundenen Treffer entspricht). Nach dem Abspielen der Sequenz setzt automatisch das Abspielen des folgenden Treffers ein. Die Abfolge des Abspielens entspricht der Reihung der Ergebnisse in der Trefferliste. Das Videofenster verfügt über eine interaktive Zeitachse, auf der die aktuelle Position in der Videosequenz und auch deren weitere Ergebnisse abgebildet werden.

Erweiterung der Indexierungstechnologie und ihre Verwendung in anderen Archiven

Bei der Präsentation der schrittweise innovierten Versionen der Such-Software haben sich an uns im Lauf der Jahre weitere Institutionen, die über umfangreiche Archivdaten verfügen, mit dem Ersuchen um deren automatische Bearbeitung gewendet. Die erste dieser Institutionen war das Institut für das Studium totalitärer Regime (ÚSTR, tschech.: Ústav pro studium totalitních režimů), das überwiegend bilanzierende Gespräche mit Zeitzeugen der Totalitarismen des 20. Jahrhunderts archiviert – d. i. Aufnahmen, die formal mit dem

Material des USC Shoah Foundation Visual History Archive praktisch ident sind. Darüberhinaus enthält das Archiv des ÚSTR auch eingescannte urkundliche Dokumente, welche die Mitarbeiter des Instituts ebenso zugänglich machen wollten. Wir haben also ein gemeinsames Projekt mit der Bezeichnung *System für eine dauerhafte Aufbewahrung der Dokumentation und Präsentation historischer Quellen aus der Periode der totalitären Regime erarbeitet* (tschech.: *Systém pro trvalé uchování dokumentace a prezentaci historických pramenů z období totalitních režimů*). (Projekt des Kulturministeriums der Tschechischen Republik mit der Nummer DG16P02B048 – 2016–2019).

Im Verlauf der Projektbearbeitung kamen wir zur Erkenntnis, dass man die eingescannten Dokumente als weiteren „Datenstrom“ verstehen kann, zu dessen Indexierung sich im Prinzip dieselbe Vorgangsweise anwenden lässt, wie bei der Indexierung der dekodierten Aufzeichnung von gesprochener Sprache. Wir müssen einzig in der Lage sein, die Bildinformation in einen elektronischen Textstrom umzuwandeln. Das kann selbstverständlich mittels der Technologie der optischen Zeichenerkennung (*optical character recognition* – OCR) durchgeführt werden. Dann können wir durch die in Sektion 2.2.1 beschriebene Methode einen Index aus eingescannten Dokumenten generieren – nur mit dem Unterschied, dass wir anstatt der Information über die Anfangs- und Endzeit der einzelnen Wörter nun die Information über die Koordinaten der Wörter im Rahmen der Seite in einem Index abspeichern. In der innovierten Suchschnittstelle können wir nun in beiden Indices gleichzeitig suchen.

In einem weiteren Projekt arbeiten wir dann mit dem Tschechischen Fernsehen zusammen, wobei es die Aufgabe der entwickelten Software ist, in den Aufzeichnungen der Sendung „Ereignisse“ (tschech.: *Události*) ein eingegebenes Wort aufzufinden, und zwar nicht nur, wenn es gesprochen, sondern auch wenn es als Text am Bildschirm abgebildet wurde. Insofern es sich um einen Personennamen handelt, sollte ferner auch das Vorkommen der entsprechenden Person in der Videoaufzeichnung gefunden werden. Wir haben in diesem Fall 3 Indices:

- Sprachindex
- „Schriftindex“
- Gesichtsinde

Als grundlegende Indexierungsweise kann man wieder dieselbe Methode verwenden, es ist bloß notwendig, geeignete Technologien für die Übertragung in die Textform zu benutzen. Im Fall einer Textlesung vom Bildschirm handelt es sich um die im Englischen als „*reading text in the wild*“ bezeichnete Technik, im letzten Fall geht es dann um die relativ weit bekannte Technologie der Gesichtserkennung (*face recognition*). Ein Beispiel für die Suchschnittstelle ist in Abbildung 2 dargestellt.

Abb. 2, Seite 128

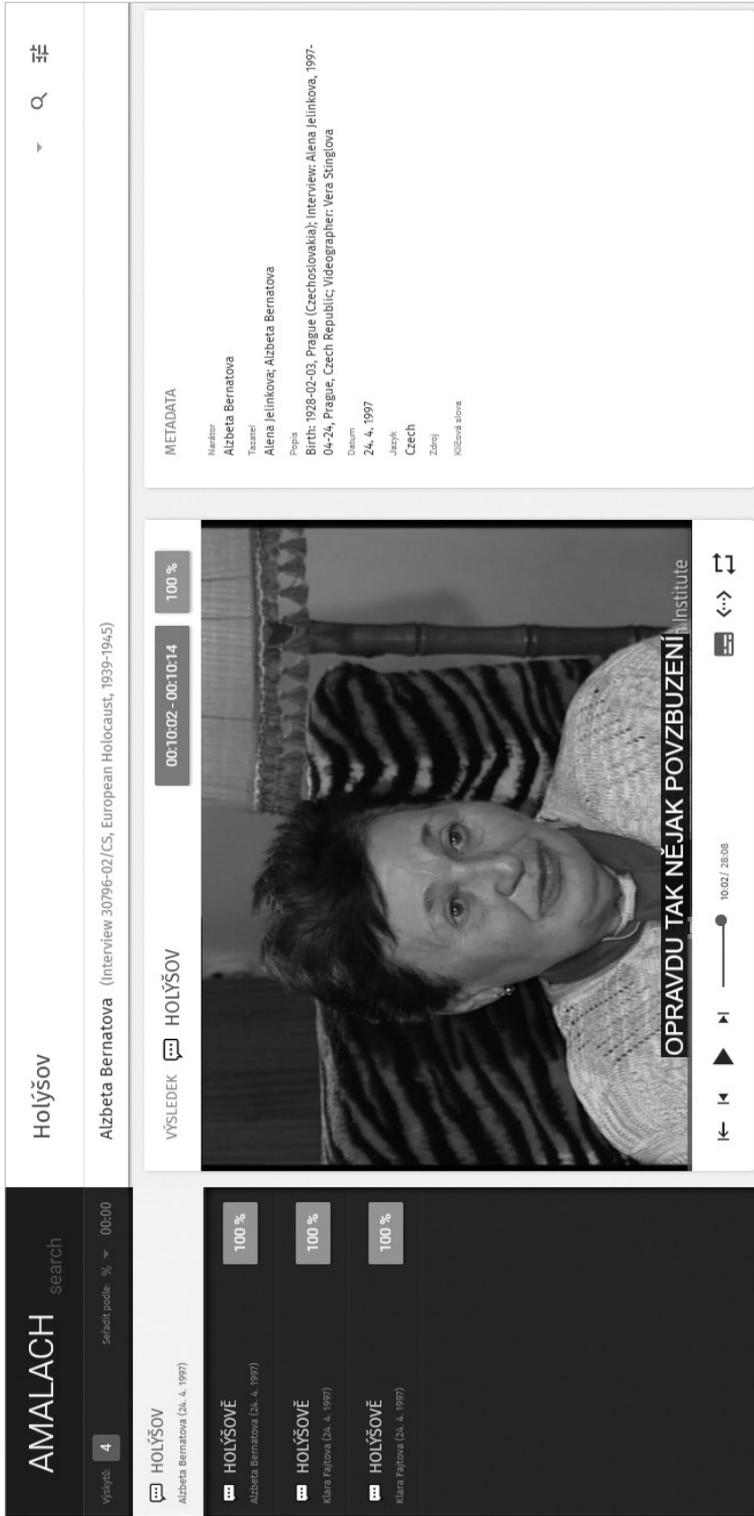


Abb. 1: Beispiel für die Suchschnittstelle

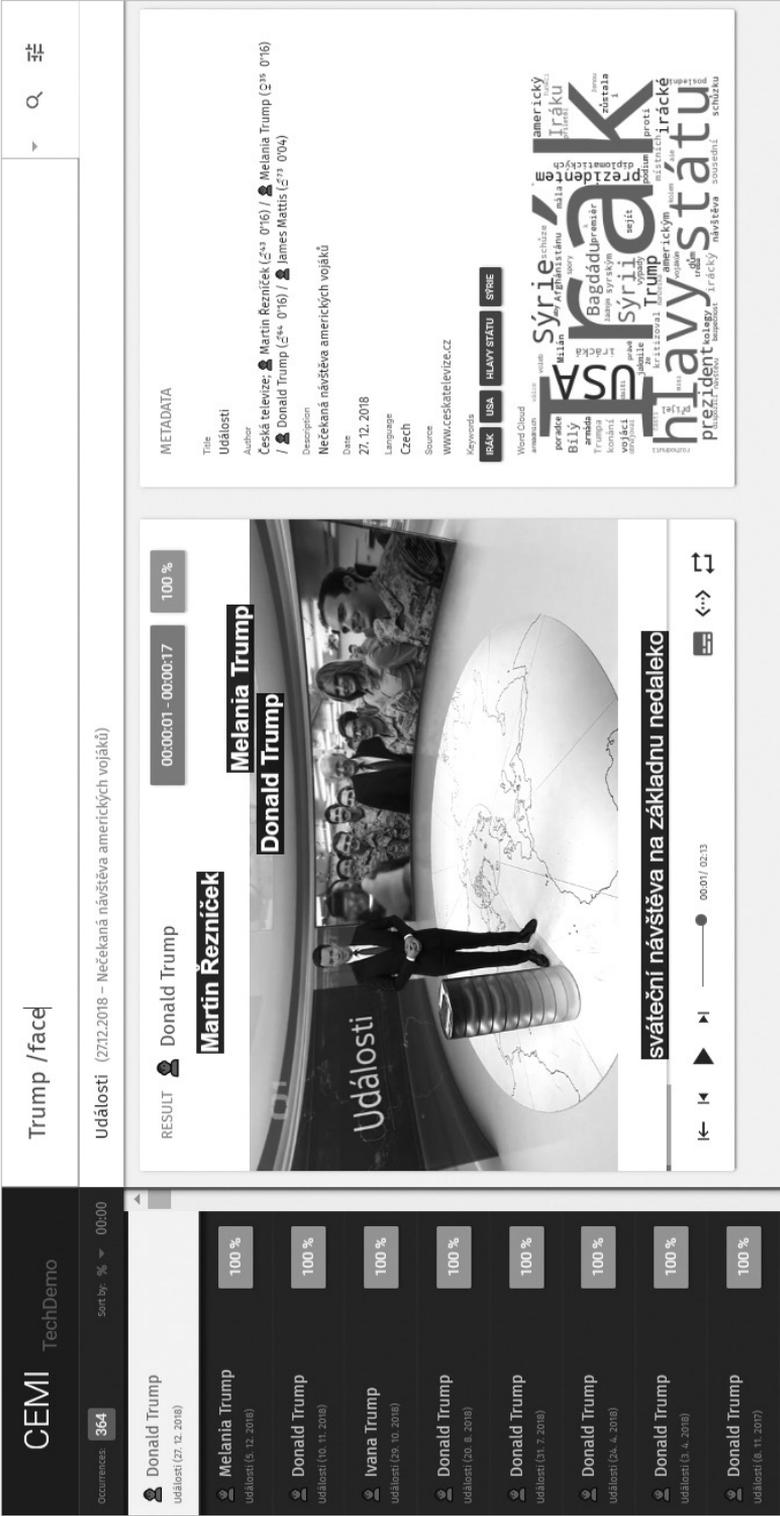


Abb. 2: Beispiel für die Suchschnittstelle

SERVICE DESIGN FÜR DEN DIGITALEN LESESAAL.¹

Ein Werkstattbericht der Staatsarchive Basel-Stadt und St. Gallen

Esther Baur, Lambert Kansy, Martin Lüthi

Einleitung

Der vorliegende Beitrag stellt den Service Design-Prozess vor, der im Rahmen des Projektes digitalAccess2archives durchgeführt worden ist. Es handelt sich um ein gemeinsames Projekt der Staatsarchive Basel-Stadt und St. Gallen zur Schaffung eines digitalen Lesesaals, der die notwendigen Funktionen umfasst, sowohl analoges wie auch digitales Archivgut zugänglich zu machen. Ziel ist die Schaffung einer nachhaltig betreibbaren Lösung mit sehr hoher Usability, durchgängiger Prozessabwicklung ohne Medienbrüche und eine Entlastung der archivinternen Abläufe durch Automatisierung.² Im Rahmen der Erarbeitung der Anforderungen an die Lösung wurde von 2016 bis 2018 ein Service Design-Prozess durchgeführt, mit dem die Nutzerperspektive systematisch in das Projekt einbezogen wurde, die bestehenden nutzungsbezogenen Services analysiert und weiterentwickelt wurden und schliesslich ein Service Design-Konzept entwickelt wurde, das über das Projekt digitalAccess2archives hinausreicht. Nachstehend werden die Erfahrungen bei der Konzipierung und Durchführung dieses Prozesses sowie seine wesentlichen Ergebnisse vorgestellt.

Die Entscheidung, in der Konzeptphase die Anforderungen und Bedürfnisse der Benutzerinnen und Benutzer vertieft einzubeziehen, wurde getroffen,

-
- 1 Der Beitrag wurde an der Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen 2019 in Prag präsentiert. Die Vortragsform wurde weitgehend beibehalten; die Präsentation ist zugänglich unter der folgenden URL: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/erschlie%C3%9Fung-und-nutzung/15_Kansy_auds2019_da2a_praesentation_ohneKommentare.pdf [Abruf 16.06.2019]
 - 2 Über das Projekt wurde zuletzt auf der Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen 2015 berichtet; die Präsentation ist zugänglich unter der folgenden URL: https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/auds-2015/digitaler-lesesaal,-nutzung-digital-archivierter-daten/02-_Kansy,%20Lambert%20&%20L%C3%BCthi,%20Martin_%20digitalAccess2archives%20-%20Werkstattbericht%20Digitaler%20Lesesaal.%20Ein%20Projekt%20der%20Staatsarchive%20St.%20Gallen%20und%20Basel-.pdf [Abruf 16.06.2019]. Vgl. auch Lüthi, Martin, Kansy, Lambert: digitalAccess2archives: Werkstattbericht digitaler Lesesaal. Ein Projekt der Staatsarchive St.Gallen und Basel-Stadt. In: Mitteilungen des Österreichischen Staatsarchivs, 59 (2016), S. 109–123.

nachdem ein Widerspruch in der bisherigen Projektarbeit festgestellt wurde: einerseits wurde der Anspruch, die Nutzerinnen und Nutzer und ihre Bedürfnisse und nicht die Binnensicht des Archivs bei der Gestaltung des Digitalen Lesesaals in den Fokus zu stellen, von Anfang an erhoben. Andererseits waren 2015 eben diese Bedürfnisse und Anforderungen unserer Benutzerinnen und Benutzer noch immer nicht bekannt. Es stellte sich die Frage, wie die angestrebte hohe Usability und eine Orientierung an den Nutzerbedürfnissen erreicht werden konnten. Dabei sollten zugleich die archivischen Anforderungen und technischen Rahmenbedingungen nicht ausser Acht gelassen werden.

Service Design

Grund für den Entscheid, einen Service Design-Prozess durchzuführen und nicht etwa dem Ansatz des User Centered Designs zu folgen³, war die Überlegung, dass die Erhebung und der Einbezug der Benutzerbedürfnisse systematisch ausgehend von den Dienstleistungen der Archivs erfolgen muss, um das Ziel einer maximalen Usability erreichen zu können. Es ging nicht nur darum, die Oberfläche und die Benutzerführung im Digitalen Lesesaal nutzerorientiert zu entwickeln, sondern um die Analyse der bestehenden und neu zu erbringenden Dienstleistungen unabhängig von den Medien und Werkzeugen, mit denen sie erbracht werden. Schliesslich findet ein Service in vielen Fällen nicht nur online oder vor Ort statt, sondern er besteht aus einer Gemengelage von Aktivitäten in beiden Dimensionen: online und on-site. Der Lesesaal als physischer Ort muss bei der digitalen Nutzung von Archivgut keineswegs obsolet werden. Im Gegenteil wird er neue Rollen und Funktionen erhalten.

„Service-Design bezeichnet den Prozess der Gestaltung von Dienstleistungen. Es ist ein Teilgebiet des Design und wird von Designern normalerweise in enger Zusammenarbeit mit Unternehmen oder Organisationen ausgeführt, um methodisch kunden- und marktgerechte Dienstleistungen zu entwickeln.“⁴ Es handelt sich um einen relativ neuen Ansatz zur kundenorientierten Entwicklung resp. Verbesserung von Dienstleistungen und der für diese benötigten Prozesse, Interaktionskanäle und verwendeten Werkzeuge.⁵

Kernelemente des Service Design-Ansatzes sind folgende Elemente: Strategiebezug, User Centered, Co-Creation, Sequencing, Evidencing und Holistic. Im Einzelnen heisst dies, Dienstleistungen ausgehend von Strategien oder auch den rechtlichen Grundlagen der Institution zu entwickeln und dabei den User und seine Bedürfnisse ins Zentrum zu stellen. Co-Creation meint den ak-

3 User Centered Design konzentriert sich auf die Gestaltung eines Produkts, in unserem Fall auf die Gestaltung einer Software-Lösung für den Digitalen Lesesaal; vgl. den Wikipedia-Artikel zu Nutzerorientierter Gestaltung, https://de.wikipedia.org/wiki/Nutzerorientierte_Gestaltung [Abruf 16.06.2019].

4 Wikipedia-Artikel Service Design, https://de.wikipedia.org/wiki/Service_Design [Abruf 16.06.2019]

5 Vgl. zu der breit angelegten Definition von Service Design Stickdorn, Marc, Schneider, Marc (Hg.): This is Service Design Thinking. Basics – Tools – Cases, 2011 BIS Publishers, Amsterdam oder auch Thoelen, Annelies, Cleeren, Steven (Hg.): Public Service Design: A guide for the application of service design in public organisations, 2015 Design Vlaanderen

tiven Einbezug der Stakeholder oder von Usern in die Entwicklung, während Sequencing die Strukturierung und Gliederung von Services in die drei Phasen Pre-Service, During-Service und Post-Service umfasst. Mit Evidencing werden für Nutzer nicht sichtbare Aspekte von Services sichtbar gemacht; dies ist der Weg, die Organisationssicht und die internen Prozessmodellierungen einzubinden. Holistisch ist der Ansatz, insofern er Dienstleistungen unabhängig von den Werkzeugen, Medien und Kommunikationskanälen betrachtet.

Im Zuge der Arbeiten sollte zum einen vermieden werden, nur die heute bereits bestehenden Prozesse und Dienstleistungen anzuschauen und zu überprüfen und zum anderen nur bereits bestehende Nutzergruppen zu befragen und zu berücksichtigen. Es galt auch Personen, die das Archiv heute nicht nutzen, auf ihre Bedürfnisse und Erwartungen an die Dienstleistungen des Archivs zu befragen sowohl im digitalen Bereich als auch in Bezug auf die Nutzung vor Ort. Ziel war es auch, Hürden zu erkennen, die eine Nutzung des Archivguts durch potentielle Nutzerinnen und Nutzer behindern und nach Möglichkeiten zu suchen, diese abzubauen und neue Dienstleistungen zu entwickeln, die die Nutzerbedürfnisse abdecken. Schliesslich sollte auch geprüft werden, ob bestehende Dienstleistungen zwar gut funktionieren, aber auf keine Resonanz bei den Nutzerinnen und Nutzer mehr stossen. In diesem Fall galt es zu prüfen, ob sie im Digitalen Lesesaal weiterhin angeboten werden sollen oder eingestellt werden können.

Da mit dem Service Design Ansatz nicht nur digitale Dienstleistungen, sondern auch Services vor Ort in den Blick genommen werden, ergab sich für beide Archive, die einmalige Chance, den Digitalen Lesesaal zu konzipieren. Dies zu einem Zeitpunkt, an dem ein Neubau des Archivs geplant resp. der räumliche Entwicklungsbedarf intensiv diskutiert wurde und somit auch die gebauten Räumlichkeiten der beiden Archive unter Berücksichtigung der Entwicklung digitaler und online Nutzung neu gedacht werden können.⁶

Durchführung

Der Service Design-Prozess begann mit einer Analyse der rechtlichen Grundlagen und der jeweiligen Archivstrategien beider Archive in Bezug auf die für die Fragestellung der Nutzerorientierung relevanten Aussagen dieser Grundlagen. Es folgte eine Analyse der bestehenden Services durch Mitarbeitenden und mit Unterstützung eines externen Spezialisten, der den weiteren Prozess begleitete und strukturierte. Damit Nutzerinnen und Nutzer involviert werden konnten, wur-

.....

6 Für das Staatsarchiv Basel-Stadt und das Naturhistorische Museum Basel wurde 2015 ein Architekturwettbewerb für einen Neubau abgeschlossen (vgl. <https://www.hochbauamt.bs.ch/wettbewerbe/abgeschlossene-wettbewerbe/naturhistorisches-museum-und-staatsarchiv-Basel-Stadt-neubau.html> [Abruf 16.06.2019]) und das Siegerprojekt bis 2018 ausgearbeitet (vgl. http://www.grosserrat.bs.ch/de/geschaeftedokumente/datenbank?such_kategorie=1&content_detail=200108974 [Abruf 16.06.2019]). Hinsichtlich des Raumbedarfs des Staatsarchivs St.Gallen wurde 2016 von einer Vorberatenden Kommission des Kantonsrats der Bericht des Regierungsrates zu dieser Thematik behandelt vgl. <https://www.ratsinfo.sg.ch/content/ris/home/geschaeftedokumente/geschaeftssuche.geschaeftdetail.html?geschaeftid=F9B85720-311C-465D-8ADA-DC226E1F1A5E&ziel=1> [Abruf 16.06.2019])

de parallel dazu eine Stakeholderanalyse durchgeführt und die Gesamtheit der Benutzenden ausdifferenziert in unterschiedlich definierte Benutzergruppen mit spezifischen Anliegen, Voraussetzungen, Bedürfnissen und Erwartungen. Eine Betrachtung der bestehenden Dienstleistungen durch Nutzervertreter schloss sich daran an, um die bisherige User Experience herauszuarbeiten und dies zu verbinden mit der Erarbeitung von Erwartungen und Bedürfnissen der unterschiedlichen Nutzergruppen an künftige Services (Co-Creation). Auf dieser Basis der internen und externen Serviceanalyse konnten dieselben beurteilt, Veränderungsbedarf eruiert und Verbesserungspotentiale aufgezeigt werden. Ebenfalls Teil dieser Phase waren Überlegungen zu einem künftigen Service-Portfolio. Die Ergebnisse aller Schritte wurden jeweils ausführlich dokumentiert und bilden einen umfangreichen Informationsbestand, der bei der Umsetzung der abschliessend erarbeiteten Service Design-Konzeption wichtig ist.

Am Anfang des Kontakts mit dem User, dieser Entdeckungsreise, stand die interne Analyse der bestehenden Dienstleistungen inkl. Mängeln und Veränderungsbedarf sowie dem Archiv bekannten Feedback seitens der Nutzenden. Das tabellarisch intern zusammengetragene Wissen wurde anschliessend durch unseren Spezialisten zu einer Touchpoint Matrix für jeden bestehenden Service aufbereitet.

Eine Touchpoint Matrix gliedert die Dienstleistungen in drei Phasen: die Aktivitäten vor der eigentlichen Leistungserbringung (Pre-Service), die Schritte, die den Kern der Leistungserbringung ausmachen (During Service) und die Arbeiten, die nachher stattfinden (Post-Service). In der Darstellung verläuft der Service von links nach rechts und gliedert sich in eine zentrale horizontale Linie, auf der die jeweiligen Berührungspunkte der Benutzenden mit dem Archiv abgebildet werden. Oberhalb dieser Touchpoints werden die Aktivitäten der Benutzenden im Verlauf der Serviceerbringung abgebildet, unterhalb werden die archivinternen Abläufe mit ihren Akteuren dargestellt. Auch die jeweils verfügbaren und eingesetzten Werkzeuge, medialen Kanäle werden sichtbar gemacht. Im Gegensatz zu Prozessdarstellungsnotationen wie BPMN wird in dieser Darstellung auch sichtbar, welche Erfahrungen die Nutzenden mit dem Dienstleistungsangebot des Archivs machen können.

Abb. 1, Seite 137

Aufbauend auf Stakeholderanalysen in Basel-Stadt und St. Gallen wurden 12 Nutzergruppen identifiziert und detailliert beschrieben. Die Beschreibung umfasste folgende Aspekte: Hintergrund (wie z. B. Herkunft, Bildung), Mindset resp. Denkhaltung, Bedürfnisse, Fragestellungen, Motivation zur Archivnutzung, Art der Beziehung und Kommunikation zwischen Nutzenden und Archiv, die Affinität zu Information resp. die Informationskompetenz, die Affinität zu E-Cultures und zur Internetnutzung, die Identifikation mit dem Ort, der Stadt oder der Region sowie schliesslich die Identifikation mit dem Archiv.

Folgende Benutzergruppen wurden gebildet: Fachdidaktiker, Direktbetroffene, Kulturschaffende resp. die Kreativwirtschaft, Freizeitnutzer, Verwaltungsmitarbeitende, Wissenschaft, Medienschaffende, Migranten, Schüler, Kunst- und Designstudierende, Architekten und Bauwesen und schliesslich die Wirtschaft.

Personas repräsentieren typische und als konkrete Personen vorgestellte Vertreter einer Benutzergruppe, die erarbeitet werden, um die Benutzergruppen besser greifbar zu machen. Die Ausdifferenzierung des allgemeinen Begriffs des

Publikums in unterschiedliche Nutzergruppen zeigte rasch auf, dass eine ganze Reihe derselben heute bereits das Archiv intensiv nutzen (z. B. Verwaltung, Wissenschaft, Direktbetroffene, Medien, Studierende), während andere als Nutzer so gut wie inexistent sind (z. B. Wirtschaft, Freizeitnutzer, Migranten, Schüler). Gruppen wie Kulturschaffende oder Kreativwirtschaft stellen heute nur peripher das Archiv aufsuchende Benutzergruppen dar, denen wir aber ein Potential für intensivere Nutzung zuschreiben. Hier und bei den heutigen Nichtnutzern ist zu fragen, welche Hindernisse für eine Archivnutzung bestehen.

In einem nächsten Schritt wurden Vertreterinnen und Vertreter der Nutzergruppen zu 10 Workshops eingeladen um ihre Sicht auf die bestehenden Dienstleistungen systematisch zu erheben und ihre Erwartungen in Bezug auf die Entwicklung künftiger Dienstleistungen kennen zu lernen. Die Workshops fanden Ende 2016 und anfangs 2017 in Zürich, St. Gallen und Basel statt. Die jeweiligen Fragestellungen in den Workshops waren weit gefasst und bezogen folgende Aspekte und Fragen mit ein:

- Was ist ein Archiv, ein Staatsarchiv?
- Wie wird das Archiv wahrgenommen?
- Wird das Archiv genutzt, falls ja aus welchen Gründen oder falls nicht, welche Hindernisse bestehen?
- Wie wird die aktuelle Servicequalität beurteilt?
- Was müsste verbessert werden?
- Welche Services fehlen und werden vermisst?

Zwei Workshops befassten sich auch mit konkreten Überlegungen, wie der Zugang zu Archivgut zeitgemäss ermöglicht werden kann: Gymnasiasten sowie Kunst-/Design-Studierende entwarfen ihre jeweiligen persönlichen Visionen und Darstellungen derselben künftiger Archivdienstleistungen. So wurden auch die Zugangsformen diskutiert und zahlreiche Anregungen gesammelt, welche Art von Zugängen aus Sicht einer jüngeren Generation «zeitgemäss» wären.

Abb. 2, Seite 138

Die Feedbacks der Benutzervertreter und Verbesserungsvorschläge sowie Hinweise auf Problemstellen in den bestehenden Dienstleistungsabläufen wurden nach den Workshops gesammelt, konsolidiert und auf die Touchpoint Matrix-Darstellung übertragen. Durch die Überlagerung aller Rückmeldungen konnte rein visuell sichtbar gemacht werden, welche Dienstleistungen aus Sicht der Mehrzahl der Benutzergruppen kritisch oder zumindest unbefriedigend sind.

Die Vielzahl der Anforderungen und Präferenzen, die aus den Workshops und den Analysearbeiten resultierten, wurde in neun Dimensionen gebündelt:

- (1) Zeitfaktor: Wie rasch muss eine Dienstleistung erbracht werden, wie hoch ist die Dringlichkeit einer möglichst unmittelbaren Ausführung und ist diese punktuell oder strukturell gegeben?
- (2) Zugänglichkeit des Archivs: welche Bedeutung hat ein möglichst niederschwelliger Zugang zu den Angeboten oder können Benutzende Hürden und Hindernisse bei der Nutzung überwinden?
- (3) Umgang mit Informationskomplexität: Welche Komplexität in der Aufbereitung der Informationen über das Archiv oder des Archivguts ist für welche Benutzergruppe vertretbar?
- (4) Vollständigkeit und Qualität der Information: Wie wichtig ist die Bedeutung, Vollständigkeit und Exaktheit der Informationen im Archiv für die Benutzerinnen und Benutzer?

- (5) Digitale Orientierung der Nutzenden: handelt es sich bei den Benutzenden um digital Immigrants oder Digital Natives und wie selbstverständlich werden digitale Dienstleistungen akzeptiert?
- (6) Materialität des Archivguts: Ist für die Nutzung das physische Objekt oder die digitale Verfügbarkeit von Information von Bedeutung?
- (7) Interaktionsweisen: Wie möchten Benutzende mit dem Archiv interagieren und in Kontakt treten? Wollen sie persönlich angesprochen werden oder das Archiv anonym resp. mithilfe von Automatismen und Algorithmen nutzen. Auch die Frage nach dem Kommunikationskanal, der verwendet werden soll für die Nutzung, online oder on-site, ist Teil dieser Dimension.
- (8) Kommunikationsart: Welche Kommunikationsart wird bevorzugt: Formell und standardisiert oder informell und situativ ausgerichtet?
- (9) Selbstständigkeit: Wie selbstständig soll das Archiv genutzt werden?

Ergebnisse

Das Ergebnis des gesamten Prozesses wurde in Form von zwei abschliessenden Berichten, für das Staatsarchiv St. Gallen und das Staatsarchiv Basel-Stadt, zusammengetragen. Diese Differenzierung war notwendig, da die jeweiligen Rahmenbedingungen, wie zum Beispiel die Archivstrategien, und der Verwaltungskontext unterschiedlich sind. Die Kernaussagen sind für beide Archive aber identisch.

Es wurde eine Reihe strategischer Ziele definiert, auf die jede Arbeit im Rahmen des Service Designs abzielen sollte:

- Das Archiv bietet einer Vielzahl von Menschen vor Ort mit unterschiedlichen Bedürfnissen attraktive Dienstleistungen an.
- Das Archiv versteht sich als sozialer Raum; es ermöglicht gesellschaftliche Teilhabe und Interaktion.
- Das Archiv ist eine offene und transparente Plattform und Drehscheibe für den Umgang der verwahrten Informationen.
- Nutzer und Archivpersonal begegnen sich auf Augenhöhe.
- Die Dienstleistungen sind informativ und unterhaltsam.
- Ein umfassender Online Access ist verzahnt mit On-Site Angeboten.
- Das Archiv positioniert sich klar in der Öffentlichkeit und stärkt auf diese Weise seine Legitimation.
- Der Einsatz von Technologie soll die selbstständige Nutzung durch die Nutzenden befördern.
- Das Archiv wandelt sich in eine agile und lernende Organisation.

In einem weiteren Schritt wurden sieben Nutzungscluster definiert. Diese gruppieren die unterschiedlichen Nutzungslagen und -interessen derart, dass eine überschaubare Menge von Feldern entsteht, die das Archiv bedienen kann.

- Professionell Wissen generieren und Information bearbeiten.
- Private Interessen verfolgen, rechtliche Ansprüche geltend machen.
- Lernen und sich weiterentwickeln.
- Kulturelle Inhalte schaffen und medial vermitteln.
- Sich unterhalten lassen.
- Vor Ort verweilen.
- Sich vernetzen.

Abb. 3, Seite 139

Diese Cluster sind nicht für alle Personas gleichermassen relevant und auch in Bezug auf ihre Medialität unterschiedlich definiert: die einen beinhalten sowohl online wie on-site Services, andere nur online oder on-site Leistungen. Die Clusterbildung zeigt auf, welche Qualitäten die einzelnen Services bieten müssen, sollen sie von den Nutzergruppen positiv wahrgenommen werden. Auf dieser Basis kann die bestehende Servicelandschaft weiterentwickelt werden.

Die Bedürfnisse der Benutzerinnen und Benutzer reichen deutlich über die klassische Sicht auf Benutzergruppen hinaus: nicht nur Wissenschaft und Verwaltung sind relevante Gruppen, die das Archiv nutzen. Andererseits besteht die Nutzerschaft des Archivs keineswegs aus der nicht weiter differenzierbaren «allgemeinen Öffentlichkeit», sondern aus einer Reihe klar unterscheidbarer Nutzergruppen mit spezifischen Bedürfnissen. Diese werden durch das definierte Set an Nutzungslagen resp. den diesen zugewiesenen und auf diese abgestimmten Services befriedigt. Die Nutzungscluster sind in digitaler wie analoger Ausprägung gedacht. Lediglich die Schwerpunkte sind unterschiedlich. Die Umsetzung dieser Konzeption und die Veränderung bestehender resp. Etablierung neuer Service erfolgt in unterschiedlichen Umsetzungsprojekten. Hierzu zählen der Digitale Lesesaal als gemeinsames Projekt der Staatsarchive Basel-Stadt und St. Gallen und – für das StABS – das gemeinsame Neubauprojekt mit dem Naturhistorischen Museum.

Beurteilung

Der geschilderte Service Design-Prozess wurde 2018 abgeschlossen. Er lässt sich folgendermassen bilanzieren:

Der Prozess erforderte viel Zeit und personellen Einsatz; deutlich mehr als unter dem Stichwort Usability in der ursprünglichen Projektplanung des Digitalen Lesesaal-Projektes vorgesehen war. Ausgehend von dem Ziel der Nutzerorientierung wandelte sich der Blick und betrachtete die Dienstleistungen, die mit der Nutzung von Archivgut zu tun haben, ganzheitlich. Die Durchführung hat deutlich gemacht, dass die Realisierung des Digitalen Lesesaals von der Archivstrategie, dem Umfeld des Archivs und den gesellschaftlichen wie technologischen Trends ausgehen muss.

Das Ergebnis zeigt, dass durch diese Arbeit eine neue Perspektive auf den Digitalen Lesesaal möglich ist; die Anforderungen, die dieser erfüllen muss, soll er den unterschiedlichen Benutzerlagen gerecht werden können, konnten geklärt werden. Vernetzung, Automatisierung und online Angebote auf der einen Seite, direkter persönlicher Kontakt und die Benutzung der analogen Unterlagen in Originalform auf der anderen Seite sind beides zentrale Anforderungen fast aller Personas und in zahlreichen Nutzungslagen gefordert. Der Digitale Lesesaal alleine kann diese Anforderungen nicht einlösen. Er muss daher zwingend gemeinsam mit on-site Lösungen und Angeboten betrachtet werden. Es geht nicht nur darum, ein effizientes und positiv wahrgenommenes Benutzererlebnis im Digitalen Lesesaal zu schaffen, sondern die Bedürfnisse der Benutzerinnen und Benutzer ganzheitlich zu bedienen.

Wir sind bei der Durchführung des Service Design-Prozesses in Kontakt getreten mit einer ganzen Reihe von bestehenden und potenziellen Nutzergruppen. Es hierbei wurden viele, teilweise ganz konkrete, Informationen zusam-

mengetragen, die sich insbesondere auf die Entwicklung und Verbesserung von online Angeboten beziehen. Zugleich haben wir auch Anregungen erhalten, wie online Angebote mit on-site Angeboten kombiniert und ergänzt werden können. Aufgrund dieser ermutigenden Erfahrungen lässt sich eine Fortsetzung des Einbezugs der Benutzerinnen und Benutzer bei der Realisierung des Digitalen Lesesaals vorstellen.

Ausblick

Unmittelbar nutzbar gemacht wurden die Erkenntnisse aus dem Service Design, insbesondere die Nutzungs-Cluster, bereits bei der Erarbeitung des Konzepts für die Informationsarchitektur des digitalen Lesesaals. Wir konnten nun die Benutzerinnen und Benutzer viel differenzierter ins Zentrum stellen, als dies sonst möglich gewesen wäre. Die erhobenen Anforderungen der Nutzenden fliessen in den Anforderungskatalog ein. Neben den funktionalen und nicht-funktionalen Anforderungen entsteht ein Design Briefing, um sicherzustellen, dass die Lösung serviceorientiert realisiert werden kann.

Was ist ein Staatsarchiv? Diese Frage wurde in allen Workshops gestellt und durch die Teilnehmer vollkommen unterschiedlich beantwortet. Das Statement «Ein Staatsarchiv ist eine Tüte voller Wunder» stammt aus einem der Workshops und vermittelt ein überraschendes Bild vom Archiv. Zugleich enthält es aber auch eine Aufforderung an uns Archivarinnen und Archivare: diese Wunder, das Archivgut, den Benutzerinnen und Benutzern zugänglich und nutzbar zu machen. Aus Sicht der Benutzenden steht nicht das Archiv im Zentrum, sondern das Archivgut, das sie nutzen möchten. Bei der Entwicklung und Gestaltung unserer Dienstleistungen sollte darüber hinaus stets gegenwärtig sein, dass im Fokus die Benutzerinnen und Benutzer stehen sollten, nicht das Archiv und seine Mitarbeitenden.

Summary

This article presents the service design process carried out as part of the digitalAccess2archives project. It is a joint project of the State Archives of Basel-Stadt and St.Gallen to create a digital reading room with the necessary functions to make both analogue and digital archive material accessible. The aim is to create a solution that can be operated in a sustainable manner with very high usability, consistent process handling without media breaks and a relief of internal archive processes through automation. As part of the development of the requirements for the solution, a service design process was carried out from 2016 to 2018 to integrate the user perspective systematically and methodically into the project. The existing access services were analysed and finally a service design concept was developed. The experience gained in the design and implementation of this process and its main results are presented in this article. These include a strategic orientation of the archive towards the needs of its users concerning the design of its access services as well as the insight that, in addition to online services, on-site services will also be needed in the future.

4. Recherche (Online, Onsite)

5. Registrierung und Verwaltung Benutzende

SIABS Service Design UX Analyse: Interne Sicht

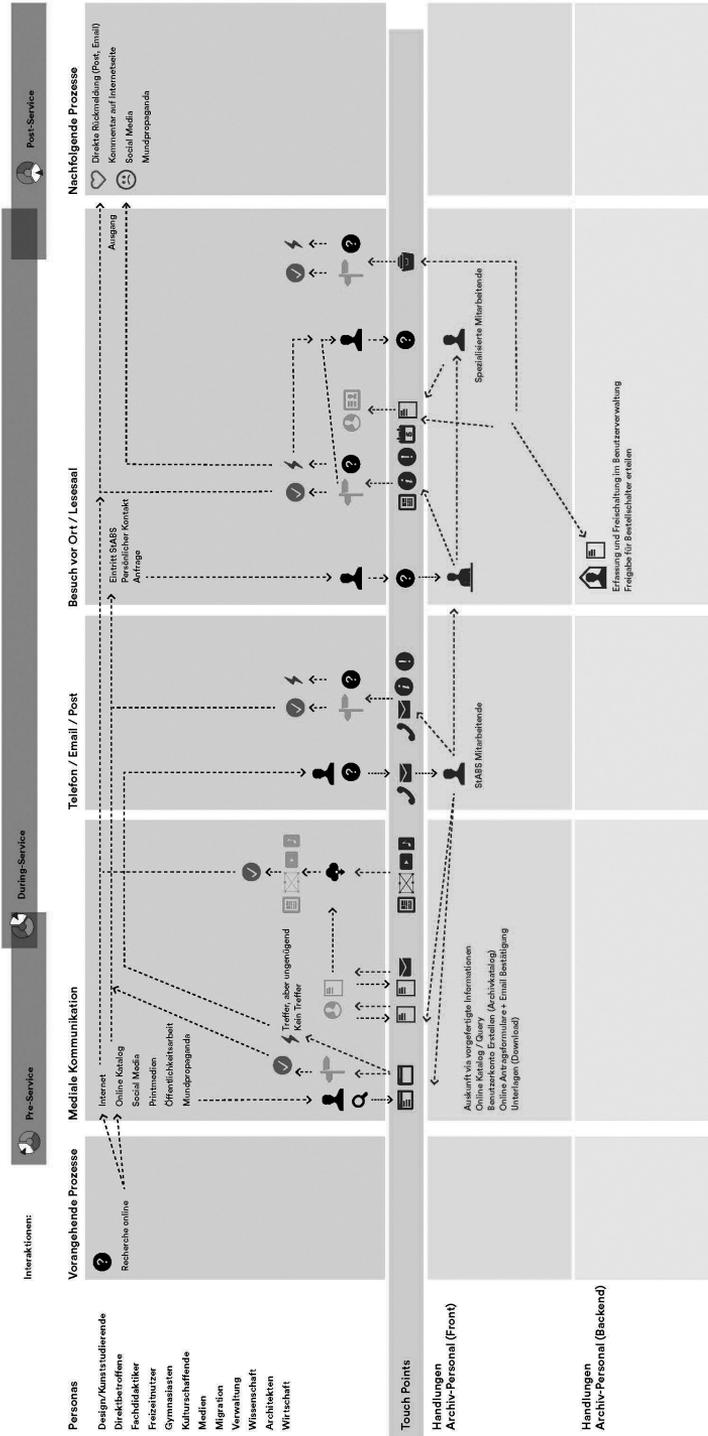


Abb. 1: Touchpoint-Matrix eines Service; Quelle: Interne Analyse Services StABS 2017

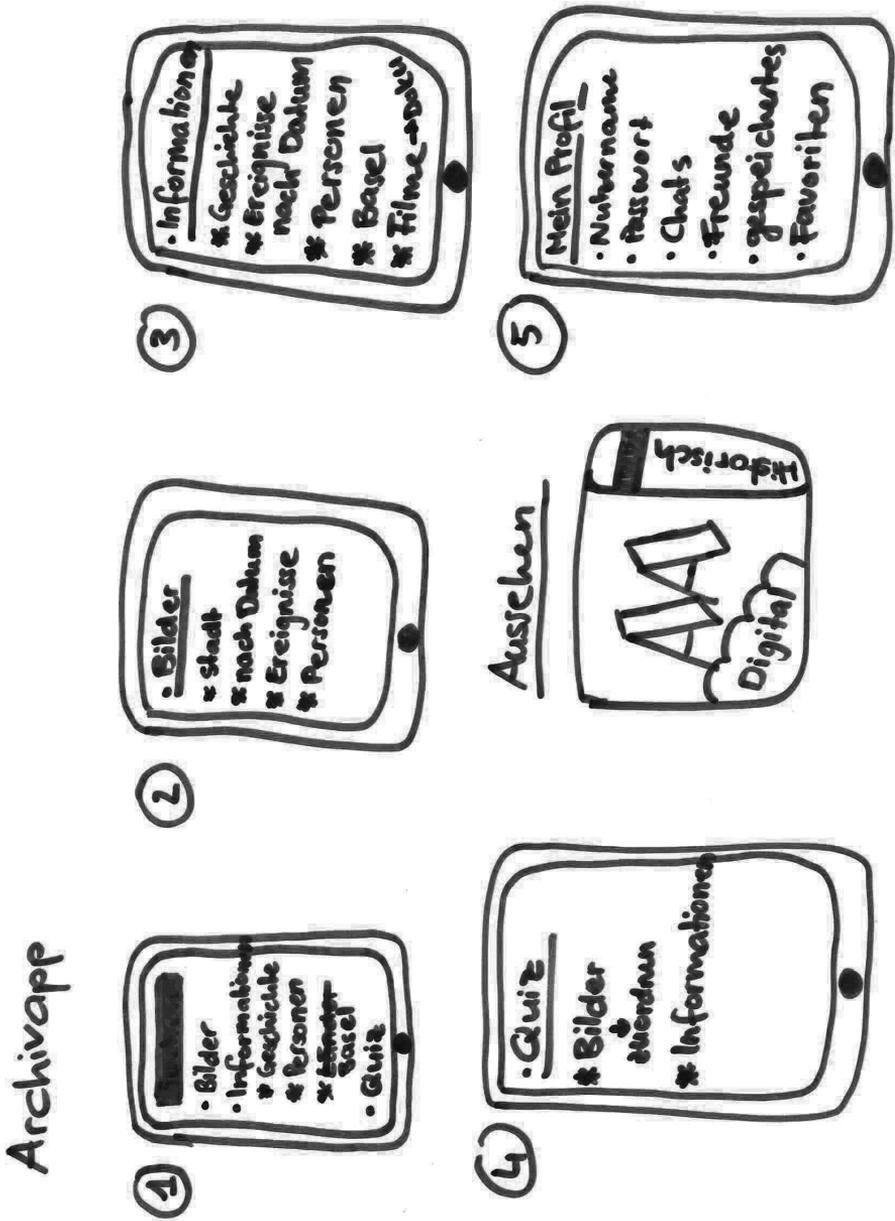


Abb. 2: Beispiel für Co-Creation-Resultate; Quelle: Fotodokumentation Service Design, Persona- Workshops 2018

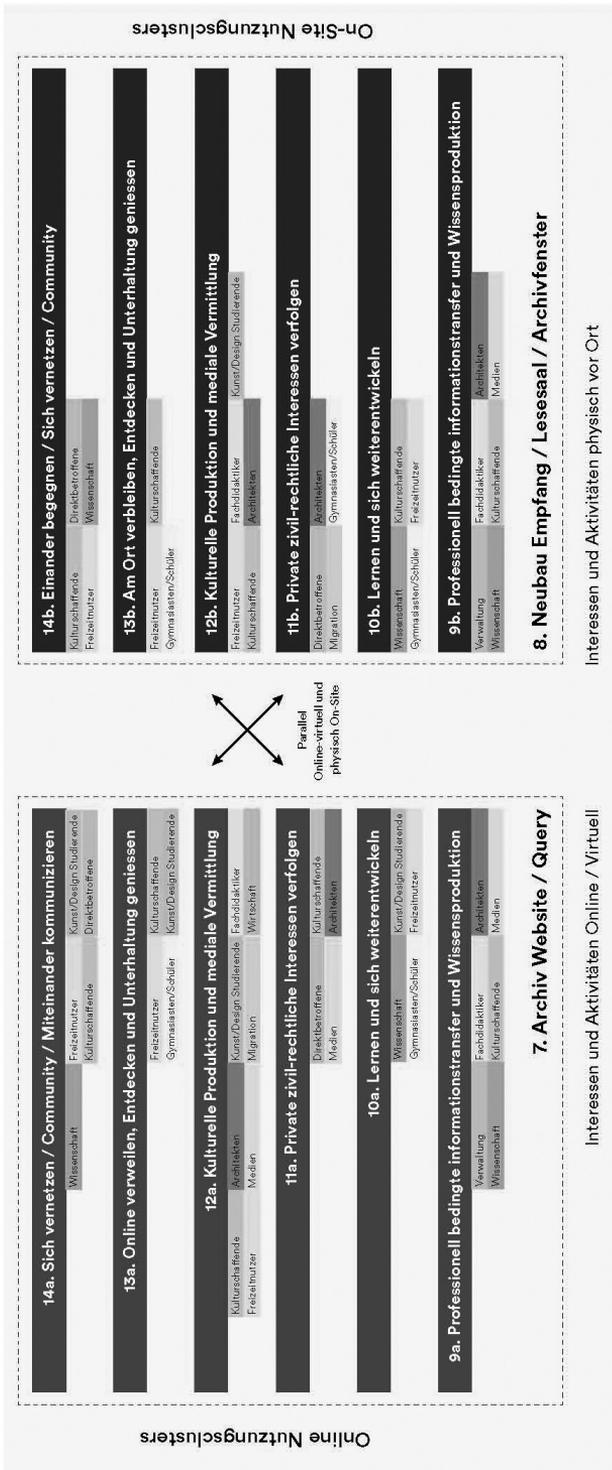


Abb. 3: Nutzungscluster und Nutzergruppen; Quelle: Service Design-Strategie StABS 2018

AUDS-CAMP

WIE GEHEN WIR
MIT DIGITALEN MEDIEN
ALS BEILAGE
EINER PAPIERNEH
AKTE UM?

WIE GEHEN WIR MIT DIGITALEN MEDIEN ALS BEILAGE EINER PAPIERNEN AKTE UM?

Zusammenfassung

Kai Naumann

Das AUdS-Camp A handelte von Datenträgerbeilagen in Papierakten. Kai Naumann schilderte zunächst die Lage im Landesarchiv Baden-Württemberg. Dort werden in einigen Bereichen, insbesondere der Strafrechtspflege, Akten schon fünf bis zehn Jahre nach Entstehung von den Staatsanwaltschaften übernommen. Ende der 2000er Jahre fiel bei Verpackungs- und Katalogisierungsarbeiten zum ersten Mal auf, dass in den Papierakten immer häufiger audiovisuelle und digitale Medien wie Audiokassetten, VHS-Kassetten, CD-ROM, DVD oder Musik-CDs als Beilagen enthalten waren. Bei solchen Objekten ist eine dauerhafte Erhaltung des physischen Trägers denkbar, aber wegen der vielfältigen Materialien und ihrer Zerfallsprozesse kaum belegbar. Ein robusterer Weg besteht darin, die Inhalte des Trägers auszulesen und in dem digitalen Archivsystem DIMAG einzulagern. In mehreren Projekten hat das Staatsarchiv einen Arbeitsablauf und Formblätter entwickelt, um diese Aufgabe zusätzlich zu den bisherigen Pflichten zu bewältigen. Die anfängliche Sorge, dass die neue Aufgabe unverhältnismäßig viel Arbeitskraft binden würde, konnte nicht bestätigt werden. Wenn eine Vollzeitkraft sich konzentriert der Aufbereitung der Datenträger widmet, können 100 bis 200 Medien pro Woche verarbeitet werden. Da der Neuzugang nur 50 bis 100 Medien pro Jahr beträgt, ist diese Aufgabe zu bewältigen.

In der Diskussion wurden ähnliche Erlebnisse mit digitalen Beilagen von anderen Tagungsteilnehmern bestätigt. Allgemein werden die Datenträger in den Akten belassen, also keine aktive Bestandserhaltung betrieben. Nur bei besonders wertvoll eingestuften Unterlagen (prominente Literaten, Politiker oder Wissenschaftler) werden die digitalen Beilagen ausgelesen und in geeigneten Archivspeichersystemen vorgehalten. Schwer haben es Archive, denen die Behörde erst nach vielen Jahrzehnten die Unterlagen überweist, denn in diesem Zeitraum können Daten- und Medienträger bereits stark zerfallen und unbenutzbar sein.

Zum Schluss stellte Naumann die in Ludwigsburg entwickelten Formblätter und Anleitungen vor und schilderte die erforderlichen Schritte im Einzelnen. Es wurde Wert darauf gelegt, dass die Versorgung der digitalen Trägerobjekte asynchron von der sonstigen Verpackung und Katalogisierung erfolgen kann, um diese Prozesse nicht aufzuhalten. Aufgefundene Objekte erhalten einen Laufzettel, in der Papierakte wird ein Stellvertreterzettel eingelegt. Die Verpackung des Datenträgers verbleibt in der Akte. Die unter Umständen wichtige

Beschriftung eines Datenträgers wird durch ein Foto oder Scan erhalten. Im Findmittel sollen die digitalen Beilagen nicht separat, sondern unmittelbar im Kontext mit der Papierakte, auf die sie sich beziehen, dargestellt werden.

AUDS-CAMP

ELEKTRONISCH SEI DIE
AKTE, HILFREICH
UND GUT. BERATUNG
VON BEHÖRDEN BEIM
UMSTIEG AUF DIE E-AKTE

ELEKTRONISCH SEI DIE AKTE, HILFREICH UND GUT.

Beratung von Behörden beim Umstieg auf die E-Akte

Zusammenfassung

Christine Friederich, Martin Schlemmer

Je digitaler Behörden werden, desto größer wird der Beratungsbedarf. Insbesondere beim Thema E-Akte bietet sich für die Archive eine gute Gelegenheit, ihre Kompetenz in der digitalen Schriftgutverwaltung einzubringen und den digitalen Wandel in den Behörden mitzugestalten. Wie Behördenberatung gelingen kann, welche Herausforderungen es gibt und welche neuen Ideen und Impulse sich finden lassen, war Thema der hier zu besprechenden Sektion des AÜS-Camps.

In einem kurzen Impulsreferat stellten Dr. Christine Friederich und Dr. Martin Schlemmer thesenartig die Erfahrungen aus inzwischen gut zwei Jahren Behördenberatung des Landesarchivs Nordrhein-Westfalen rund um das Thema E-Government vor. Ausgangspunkt war das einem antiken griechischen Motto („Erkenne Dich selbst“) nachempfundene „Kenne dich selbst!“. Die genaue Definition der eigenen Beratungskompetenzen, so die Vortragenden, ist der erste wichtige Schritt auf dem Weg zur erfolgreichen Behördenberatung. Als weitere Erfolgsfaktoren nannten die Vortragenden eine konstruktive Beratungssituation, welche die durch den Veränderungsprozess freiwerdenden Potenziale in den Mittelpunkt stellt, die Schaffung niedrigschwelliger Beratungsangebote und den Abbau von Vorurteilen gegenüber (digitaler) Aktenführung. Herausforderungen sind die Anforderung, Beratungsangebote permanent weiterzuentwickeln und Beratungsbedarfe zu antizipieren, der Umgang mit schwierigen Beratungssituationen und auch die Kalkulation der Beratungsbedarfe, also zu welchem Zeitpunkt in welchem Umfang Beratung abgerufen wird. Damit eng verbunden ist die Frage, inwiefern diese Beratungsleistungen seitens des beratenden Archivs mit Blick auf die eigenen Ressourcen auch wahrgenommen werden können. Abschließend wurde noch kurz auf einige zielgruppenspezifische Beratungsformate eingegangen.

Im Anschluss entwickelte sich eine sehr intensive und offene Diskussion, welche die folgenden drei Aspekte in den Mittelpunkt stellte:

Wie lassen sich Führungskräfte erfolgreich einbinden?

Führungskräfte für die digitale Schriftgutverwaltung zu gewinnen, wurde auch hier in der Diskussion als ganz entscheidend bewertet. Dennoch gibt es immer wieder Probleme, überhaupt Zugang zu Führungskräften zu finden bzw. diese dann auch von der Wichtigkeit des Themas Schriftgutverwaltung zu überzeugen. Die geschilderten Erfahrungen waren sehr unterschiedlich. Während ein Archiv damit zu kämpfen hat, dass bei den Behörden digitale Verwaltungsarbeit als reines IT-Thema verstanden wird und das Archiv somit ausgeschlossen wird, haben viele andere die Erfahrung gemacht, dass einzelne Behörden oder Organisationseinheiten der E-Akte sehr positiv gegenüberstehen und dann grundsätzlich auch offen für Beratung sind. Gelingt die Beratung dort, bedeutet dies allerdings nicht zwangsläufig, dass der Funke auf andere Organisationseinheiten überspringt. Manchmal erweisen sich auch äußere Einflüsse als Türöffner. So wurde berichtet, dass mit der DSGVO auch das Thema ordnungsgemäße Dokumentation in den Fokus rückte und die Kompetenz des Archivs nachgefragt wurde. Unterstützend, so eine andere Erfahrung, wirkt sich auch der Verweis auf bereits erprobte Konzepte aus, z.B. den Führungskräfte-Leitfaden des Organisationskonzepts elektronische Verwaltungsarbeit.¹

Wie lässt es sich erreichen, als kompetente Ansprechpartnerin wahrgenommen zu werden?

Hier verwiesen insbesondere die Vertreterinnen und Vertreter der Unternehmensarchive darauf, dass das Thema DMS gerade in Unternehmen sehr stark verfolgt werde und es in jüngster Zeit zahlreiche Neueinführungen von DMS gegeben habe. Treiber der Entwicklung sind Erfordernisse an die Rechts- und Revisionssicherheit der Dokumentation, aber auch Kostenaspekte. Für Archive ist es wichtig, über solche Entwicklungen informiert zu sein, um dann die eigene Fachkompetenz einbringen zu können. Netzwerken und „die Ohren offen halten“ sind also wesentliche Voraussetzungen. Wenn es gelingt, den Fuß in die Tür zu bekommen, werden Archivarinnen und Archivare in der Regel auch gehört und ihre Kompetenz geschätzt.

Auch wenn die E-Akte gerade ein großes Thema ist – gibt es Tendenzen, dass die E-Akte in Zukunft von anderen Systemen abgelöst wird?

Das Thema E-Zusammenarbeit beschäftigt gerade viele Unternehmen und Behörden. Die oft sehr komfortablen Plattformen für die Zusammenarbeit bieten meist auch die Möglichkeit, Dokumente abzulegen und zu organisieren. Sie stellen somit gerade in Behörden eine Konkurrenz zur herkömmlichen Aktenführung dar. In der Diskussion wurden durchaus Tendenzen beschrieben, dass die Ent-

.....
1 https://www.verwaltung-innovativ.de/DE/E_Government/orgkonzept_everwaltung/orgkonzept_everwaltung_artikel.html [Abruf29.4.2019].

wicklung generell eher in diese Richtung gehen und möglicherweise auch einmal die E-Akte ablösen könnte. Aktuell erfordern aber die gültigen Rechtsgrundlagen (noch) die Aktenführung. In diesem Zusammenhang wurde auch die Frage nach Federführung und der Verantwortung in der Verwaltung generell gestellt.

Fazit

Es lässt sich festhalten, dass mit der Digitalisierung das Thema „Beratung“ in allen Archivsparten an Bedeutung gewonnen hat. Die Archive erkennen die Chance, die sich ihnen hier zur Mitgestaltung bietet und engagieren sich im Rahmen ihrer Möglichkeiten. Auch wenn es weiterhin Hemmnisse gibt – in der Diskussion überwog der Wunsch, sich in die Beratung einzubringen. Zudem zeigte die Diskussion, wie wichtig der Austausch ist, um die eigenen Erfahrungen und Erwartungen besser einordnen zu können. Wir können von den Erfolgsgeschichten profitieren und aus Misserfolgen lernen und gemeinsam nach neuen Lösungen und Ideen suchen. Dafür hat das AUdS-Camp einen weiteren Impuls geliefert.

AUOS-CAMP

MIT BPMN UND REST
ZU EINER KLAREN UND
OASIS-REFERENZMODELL-
KONFORMEN SOFTWARE-
ARCHITEKTUR

MIT BPMN UND REST ZU EINER KLAREN UND OAIS-REFERENZ- MODELL-KONFORMEN SOFTWARE-ARCHITEKTUR

Zusammenfassung

Frank Obermeit

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 und 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 allgemein als Workflows bezeichnet. Seit 2011 gibt es den Standard BPMN 2.0 von der Open Management Group (OMG)¹, der die grafische Modellierung von Workflows spezifiziert.

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 diese müssen verwaltet werden. Wenn davon ausgegangen wird, dass SIP und DIP in Storages, wenn auch nicht dauerhaft, aufbewahrt werden, dann sind dafür ebenfalls Archival Storages notwendig und sollten wie die Functional Entity Archival Storage betrachtet und konzipiert werden.

Mit REST-Web Services kann die Benutzung der Archival Storages gekapselt werden und die Dienste und Funktionen können in diesen implementiert werden. Mittels BPMN können diese REST-Web Services (Operationen) miteinander verknüpft werden.

Graphical User Interfaces (GUI) und die 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. Diesem Aspekt muss eine OAIS-konforme Software-Architek-

.....
1 <https://www.omg.org> [Abruf 4. 10. 2019]

tur gerecht werden. Die Dokumentation und Nachvollziehbarkeit abgearbeiteter Workflows muss natürlich gewährleistet werden.

Nachfolgend wird eine ganzheitliche Software-Architektur beschrieben, die genau diese Anforderungen berücksichtigt und erfüllt. Konzipiert wurde eine OAI-konforme Software-Architektur, deren Machbarkeit durch ein begleitendes Prototyping nachgewiesen wurde. Schlussendlich wird ein auf Docker bzw. Docker Compose² basierendes und out-of-the-box nutzbares Framework vorgestellt, welches von allen genutzt werden kann, die eine software-technische Umsetzung des OAIswReferenzmodells suchen.

Das Konzept und das Framework sind allgemeingültig, aber individuell anpassbar und nutzbar.

.....
2 <https://docs.docker.com/compose/> [Abruf 4. 10. 2019]

MIT BPMN UND REST ZU EINER OAIS-KONFORMEN SOFTWARE-ARCHITEKTUR

Frank Obermeit

Abstract – Es wird beschrieben, wie eine Open Archival Information System (OAIS)¹ [1] - konforme Software-Architektur mit Business Process Model and Notation (BPMN)² [2] und Representational State Transfer (REST)³ [3] konzipiert werden kann. Die Machbarkeit wurde mit einer Implementierung des Konzepts nachgewiesen. Entstanden ist ein Framework, bestehend aus Open Source Software, Community Software und Oracles no-cost-features unter Benutzung von Docker, einer Container Virtualisierungssoftware. Das Framework kann individuell eingesetzt werden, obwohl es out-of-the-box installiert werden kann.

Kernaussage

Die Kernaussage meines Beitrages lautet salopp formuliert: Die Archival Storages werden mit REST gekapselt, die Prozessmodellierung und Prozessverarbeitung erfolgt mit BPMN, alles wird über eine webbasierte Benutzeroberfläche dem Nutzer zur Verfügung gestellt und die vorhandene Rechteverwaltung wird genutzt.

Hinweis

Nachfolgend wird das Konzept strukturiert beschrieben. Der Beitrag schließt mit der Umsetzung eines fiktiven Ingest. Schauen Sie sich unabhängig davon den AUdS-Vortrag⁴ inklusive ausführlichem Skript⁵ an.

.....
1 <https://public.ccsds.org/pubs/650x0m2.pdf> [Abruf 4. 10. 2019]

2 <http://www.bpmn.org> [Abruf 4. 10. 2019]

3 https://en.wikipedia.org/wiki/Representational_state_transfer [Abruf 4. 10. 2019]

4 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 [Abruf 4. 10. 2019]

5 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 [Abruf 4. 10. 2019]

Einführung und Übersicht

Das OAI-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 und 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 allgemein als Workflows bezeichnet. Seit 2011 gibt es den Standard BPMN 2.0 von der Open Management Group (OMG)⁶, der die grafische Modellierung von Workflows spezifiziert.

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 diese müssen verwaltet werden. Wenn davon ausgegangen wird, dass SIP und DIP in Storages, wenn auch nicht dauerhaft, aufbewahrt werden, dann sind dafür ebenfalls Archival Storages notwendig und sollten wie die Functional Entity Archival Storage betrachtet und konzipiert werden.

Mit REST-Web Services kann die Benutzung der Archival Storages gekapselt werden und die Dienste und Funktionen können in diesen implementiert werden. Mittels BPMN können diese REST-Web Services (Operationen) miteinander verknüpft werden.

Graphical User Interfaces (GUI) und die 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. Diesem Aspekt muss eine OAI-konforme Software-Architektur gerecht werden. Die Dokumentation und Nachvollziehbarkeit abgearbeiteter Workflows muss natürlich gewährleistet werden.

Nachfolgend wird eine ganzheitliche Software-Architektur beschrieben, die genau diese Anforderungen berücksichtigt und erfüllt. Konzipiert wurde eine OAI-konforme Software-Architektur, deren Machbarkeit durch ein begleitendes Prototyping nachgewiesen wurde. Schlussendlich wird ein auf Docker bzw. Docker Compose⁷ basierendes und out-of-the-box nutzbares Framework vorgestellt, welches von allen genutzt werden kann, die eine software-technische Umsetzung des OAI-Referenzmodells suchen.

Das Konzept und das Framework sind allgemeingültig, aber individuell anpassbar und nutzbar.

6 <https://www.omg.org> [Abruf 4. 10. 2019]

7 <https://docs.docker.com/compose/> [Abruf 4. 10. 2019]

Projektverlauf

Im August 2017 wurde BPMN als möglicher Lösungsansatz für eine Workflow-Modellierung und Workflow-Steuerung identifiziert. Im vierten Quartal 2017 wurde nach Komponenten für eine ganzheitliche Verwaltung von Digitalisaten und Digitalen Objekten gesucht. Eine erste Idee für eine OAI3-konforme Software-Architektur wurde im Januar 2018 innerhalb des DIMAG-Verbundes und DAN-Verbundes (Digitale Archivierung Nord) vorgestellt. Von Februar bis Mai 2018 erfolgte die Einarbeitung in relevante Technologien und Produkte. Ein erster, das Konzept umsetzender Prototyp, wurde im Juni 2018 entwickelt, allerdings war der Aufwand für die Benutzung von RESTful-Web Services innerhalb der GUI unverhältnismäßig groß. Dies änderte sich im August 2018, da die GUI-Software in diesem Bereich vom Hersteller wesentlich verbessert wurde. Das Konzept inklusive prototypischer Umsetzung wurde im September 2018 im Landesarchiv Sachsen-Anhalt präsentiert und im November 2018 auf dem DIMAG-Entwicklertreffen ausführlich vorgestellt. Stand Mai 2019 hat das Framework einen fast produktionsreifen Status erreicht. Innerhalb von achtzehn Monaten entstanden ein Konzept und ein darauf basierendes Framework – oder anders formuliert – in achtzehn Monaten von der Idee bis zur Umsetzung.

Konzept - Details

Anforderungen

Es soll eine Software Architektur konzipiert werden, die die Umsetzung des OAI3-Referenzmodells für Digitalisate und Digitale Objekte ermöglicht. Etablierte Standards, bewährte Open-Source-Software und Community-Produkte sollen eingesetzt werden und Eigenentwicklungen vermieden werden. Das Graphical User Interface soll leicht bedienbar und jederzeit austauschbar sein. BPMN soll für die Workflow-Modellierung und Workflow-Verarbeitung eingesetzt werden. Die Machbarkeit des Konzepts soll durch ein begleitendes Prototyping nachgewiesen werden. Die Authentifikation und Autorisierung sind mit etablierten Technologien und Produkten umzusetzen. Verteilbarkeit und Skalierbarkeit sollen keine Probleme bereiten.

SIP, AIP und DIP benötigen (Archival) Storages

Der Archival Store bewahrt AIP dauerhaft auf und SIP müssen ebenfalls aufbewahrt werden, wenn auch nicht dauerhaft. Somit gelten für AIP und SIP ähnliche Anforderungen, die sich nur in der Speicheraufbewahrungsfrist unterscheiden. SIP werden während des Ingest bearbeitet und verarbeitet (vereinfacht: bearbeitet). AIP werden für den Access bearbeitet und als DIP aufbereitet. AIP werden vom Preservation Management für die dauerhafte Aufbewahrung bearbeitet. Das Datamanagement und die Administration benötigen für die Aufgabenplanung und das Erarbeiten von Empfehlungen einen indirekten Zugriff auf den Archival Storage. Für SIP und DIP gelten ähnliche Anforderungen.

UUID, URI und Web Services

Wenn SIP, AIP und DIP über (universally unique identifier (UUID)⁸ eindeutig identifiziert und über Uniform Resource Identifier (URI)⁹ lokalisiert werden, dann bieten sich REST-Web Services für eine Kapselung an. Jede Bearbeitung oder Verarbeitung wird über Web Service Operationen durchgeführt. Außerdem können Web Services versioniert und durch den Einsatz eines OpenAPI¹⁰ unterstützenden Frameworks sich selbst dokumentierend gestaltet werden, was die Archivierung der Servicedokumentationen erleichtert.

REST-Web Services and Web Server

REST-Web Services werden vorrangig über Web Server bereitgestellt, nutzen damit die Möglichkeiten von HTTP¹¹ aus und werden dann als RESTful-Web Services bezeichnet. So kann für den Access und das Load Balancing auf bewährte technische Lösungen der Web Server zurückgegriffen werden. Die Skalierung und die Vernetzung stellen somit auch kein Problem dar. Die Storages können dadurch beliebig verteilt werden und sind örtlich nicht an ein Workflow-Framework gebunden.

Authentication and Authorisation

LDAP ist ein verbreitetes Protokoll für den Zugriff auf verteilte Verzeichnisinformationen. OAuth2¹² und OpenID¹³ erweitern die technischen Möglichkeiten.

BPMN

Mit dem OMG-Standard BPMN lassen sich Workflows graphisch modellieren. Die Notationssymbole Pools, Lanes, Tasks, Sub Prozesse, Gateways und Events bieten die größtmögliche Flexibilität bei der Modellierung. Pools identifizieren unterschiedliche Organisationen. Lanes beschreiben unterschiedliche Abteilungen. Tasks und Subprozesse können einmalig oder mehrfach parallel bzw. sequenziell abgearbeitet werden. Gateways beeinflussen den Prozessverlauf. 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 Steuerung des Prozesses, sondern können auch Process Meta Data sein.

Hinter Business Rule Tasks verbergen sich Decision Model and Notation (DMN)¹⁴ – Modelle, durch deren Einsatz sich die Programmierung von Entscheidungstabellen erübrigt.

8 https://en.wikipedia.org/wiki/Universally_unique_identifier [Abruf 4. 10. 2019]

9 https://en.wikipedia.org/wiki/Uniform_Resource_Identifier [Abruf 4. 10. 2019]

10 <https://www.openapis.org> [Abruf 4. 10. 2019]

11 https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol [Abruf 4. 10. 2019]

12 <https://oauth.net/2/> [Abruf 4. 10. 2019]

13 <https://openid.net> [Abruf 4. 10. 2019]

14 <https://www.omg.org/spec/DMN/About-DMN/> [Abruf 4. 10. 2019]

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 selbst dokumentierend. Die Modelle können versioniert und unverändert archiviert werden.

BPM-Engine

Eine BPM-Engine kann BPMN-Modelle verarbeiten. Allerdings müssen die fachlichen BPMN-Modelle um technische Parameter und Skript Snippets erweitert werden.

BPMN und DevOps¹⁵ [4]

Das BPMN- Modell erarbeiten Fachbereich und IT gemeinsam, das sie sukzessiv in einem iterativen Prozess verfeinern. Das BPMN-Modell kann in eine BPM-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)¹⁶.

DevOps und Container Virtualization¹⁷

Software Entwicklungsprozesse lassen sich durch den Einsatz von Container Virtualisierung wesentlich vereinfachen und beanspruchen weniger Ressourcen als die Virtualisierung von Betriebssystemen. Die Isolation von Programmen wird durch die Container Virtualisierung vereinfacht. Die Skalierbarkeit und Verteilbarkeit werden ohne zusätzliche Eigenentwicklungen ermöglicht.

Graphical User Interface (GUI)

Kann der Prozess nicht vollständig automatisiert werden und es sind Nutzereingaben für die Verarbeitung und Bearbeitung der SIP, AIP und DIP notwendig, dann werden dafür User Tasks modelliert. Nutzereingaben und Nutzerentscheidungen sind von Prozessdaten und den eigentlichen Daten, also SIP, AIP und DIP, abhängig. Diese Daten müssen für den Nutzer aufbereitet werden. Mit der GUI werden Daten nur aufbereitet und deshalb ist die Handhabbarkeit das wichtigste Kriterium für deren Auswahl.

Das Konzept in wenigen Sätzen

Mit REST-Web Services kann die Bearbeitung und Verarbeitung von SIP, AIP und DIP gekapselt werden. Die Bearbeitung und Verarbeitung kann mit BPMN modelliert und prozessiert werden. Nutzerinteraktionen erfolgen über eine GUI und für die Authentifikation und die Autorisierung werden bekannte Verzeichnis-

15 <https://en.wikipedia.org/wiki/DevOps> [Abruf 4. 10. 2019]

16 <https://en.wikipedia.org/wiki/DevOps> [Abruf 4. 10. 2019]

17 [https://en.wikipedia.org/wiki/Container_\(virtualization\)](https://en.wikipedia.org/wiki/Container_(virtualization)) [Abruf 4. 10. 2019]

dienste eingesetzt. Diese übersichtliche Struktur kann für die Umsetzung einer OAIS-konformen Software-Architektur angewendet werden.



Abb. 1: Software-Architektur – 300dpi-Software-Architektur.jpg

Vom Konzept zum Framework

Das Konzept beschreibt die zusammenhängende Nutzung bekannter Technologien und Standards. Die konzeptionellen Anforderungen an mögliche Software-Produkte sind alltäglich und nicht ungewöhnlich. Daher war es nur ein „kleiner“ Schritt vom Konzept bis zu einer technischen Umsetzung. Das Konzept gibt den ordnenden Rahmen für ein Framework vor. Die Produktauswahl hängt von den eigenen Kenntnissen und Fähigkeiten ab und hier wurde sich bei den Web-Servern, der BPM-Engine für Open Source, beim Graphical User Interface für no-cost-features von Oracle und bei den Datenbanken für Open-Source und no-cost-features entschieden. Die nachfolgende Abbildung zeigt das darauf aufbauende Framework.

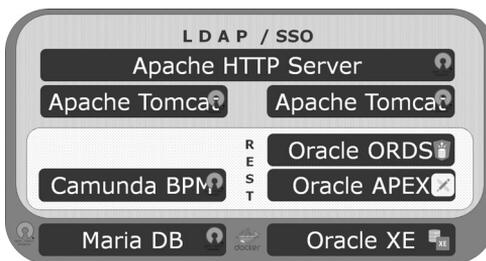


Abb. 2: Framework – 300dpi-Framework.jpg

Es wurden Oracles no-cost-features APEX¹⁸, ORDS¹⁹ und Oracle XE²⁰ eingesetzt. Als BPM-Engine wurde das Open-Source-Produkt Camunda BPM²¹ verwendet. Als Open-Source-Datenbank wurde Maria DB²² ausgewählt und als

18 <https://apex.oracle.com/en/> [Abruf 4. 10. 2019]

19 <https://www.oracle.com/database/technologies/appdev/rest.html> [Abruf 4. 10. 2019]

20 <https://www.oracle.com/database/technologies/appdev/xe.html> [Abruf 4. 10. 2019]

21 <https://camunda.com/> [Abruf 4. 10. 2019]

22 <https://mariadb.org> [Abruf 4. 10. 2019]

Web Server der Apache-http-Server²³ und die Apache-Tomcat-Software²⁴. Docker²⁵ in Verbindung mit Docker Compose wurde für die Container- Virtualisierung eingesetzt.

Das gesamte Framework kann über eine einzige Konfigurationsdatei individuell angepasst und innerhalb eines halben Arbeitstages in Betrieb genommen werden.

Die Darstellung berücksichtigt nicht die individuellen RESTful-Web Services zur Bearbeitung und Verarbeitung der SIP, AIP und DIP.

Nutzen

Entstanden ist ein Konzept für die Gestaltung eines Frameworks, mit dem das OAIS-Referenzmodell als Software umgesetzt werden kann. Es wurde ausschließlich auf bekannte und etablierte Standards, Technologien und Produkte zurückgegriffen. Eigenentwicklungen bzw. Anpassungen sind nur bei den SIP, AIP und DIP kapselnden RESTful- Web Services notwendig.

Protokollierung, Dokumentierung, Skalierbarkeit, Verteilbarkeit, Flexibilität und Benutzbarkeit sind unproblematisch. BPMN, als Standard zur Modellierung und Abarbeitung, bildet den entscheidenden Aspekt dieses Konzepts.

Der Einsatz des Frameworks ermöglicht es, sich auf das Beschreiben von fachlichen Funktionen und Diensten und die fachlichen Workflows zu konzentrieren und die technischen Anforderungen auf die Funktionsbeschreibungen der RESTful-Web Services zu beschränken. Entweder werden vorhandene Produkte gekapselt oder die Umsetzungen werden beauftragt.

Die Zusammenarbeit zwischen dem Fachbereich und der IT gestaltet sich strukturierter und konstruktiver, denn die Schnittstellen und Modelle werden sukzessiv in einem iterativen Prozess gemeinsam entwickelt und verfeinert. Projekte werden vielleicht nicht schneller, aber effektiver umgesetzt und Projektfortschritte, aber auch Fehlentwicklungen, werden eher sichtbar.

Das gesamte Framework wurde ausschließlich mit Open Source Software, Community Produkten und no-cost-features umgesetzt. Auf externes Wissen wurde nicht zurückgegriffen. Es entstanden keine zusätzlichen Kosten. Wenn ein professioneller Produkt-Support gewünscht wird, dann können die Enterprise-Versionen der oben genannten Produkte eingesetzt werden.

Umsetzung eines fiktiven Ingest-Prozesses

Ein Ingest-Prozess könnte sich aus den folgenden Arbeitsschritten (Tasks) zusammensetzen.

Ein SIP wird als BagIt²⁶ per FTP angeliefert. Darüber wird der Archivar per E-Mail informiert und parallel wird das SIP einer JHove-Validierung unterzogen.

23 <https://httpd.apache.org> [Abruf 4. 10. 2019]

24 <https://tomcat.apache.org> [Abruf 4. 10. 2019]

25 <https://www.docker.com> [Abruf 4. 10. 2019]

26 <https://en.wikipedia.org/wiki/BagIt> [Abruf 4. 10. 2019]

Über das Validierungsergebnis wird der Archivar ebenfalls per E-Mail informiert. Die E-Mail enthält zusätzlich eine URL zur webbasierten Benutzeroberfläche, wo das Validierungsergebnis visualisiert wird. Der Archivar kann dann das Validierungsergebnis bewerten und ggf. vor der SIP-AIP-Migration korrigierend eingreifen. Abschließend veranlasst der Archivar die automatische Übernahme in den Archival Storage z.B. DIMAG.

Dieser fiktive Ingest-Prozess wurde mit dem oben beschriebenen Framework innerhalb von zwei Wochen umgesetzt. Für die JHove-Validierung und die BagIt-Be- und Verarbeitung wurden zwei RESTful-Web Services implementiert, die durch den Einsatz des Open Source Framework swagger²⁷ selbstdokumentierend sind. Der gesamte Prozess besteht aus neunzehn Tasks und zwei Gateways. Die JHove-Validierung wurde als Sub-Prozess modelliert und ermöglicht eine Parallelisierung der Validierung der tatsächlichen Daten-Dateien. Eine Entscheidungstabelle (Business Rule bzw. DMN) erleichterte die flexible Umsetzung der archivspezifischen MIME-Type Policy.

Das zunächst aus fachlicher Sicht gestaltete BPMN-Modell wurde um technische Parameter ergänzt und in einem iterativen Prozess durch Archivar und Informatiker verfeinert. Zusätzliche Kommentierungsmöglichkeiten verbessern die Verständlichkeit des Modells. Die Modellierung erfolgte mit dem als Open Source Tool verfügbaren BPMN- und DMN- Modeler²⁸ der Firma Camunda²⁹. Der BPM-Engine wurde das Modell direkt vom Modeler aus übergeben.

Der entscheidende Vorteil besteht in der Verwaltung aller fachlichen und technischen Informationen in einer XML-Datei, der BPMN- bzw. DMN-Datei. Die Modelle sind damit selbstdokumentierend und liegen gleichzeitig in dem Langzeitarchivformat XML vor. Die webbasierte Benutzeroberfläche wurde mit dem no-cost-Feature APEX von Oracle intuitiv entwickelt und gestaltet. Die während der BPMN-Prozessverarbeitung anfallenden Daten und Metadaten wurden innerhalb von Script-Tasks mit einfachen Programmierkenntnissen verarbeitet und bearbeitet.

Konkret besteht das BPMN-Modell des fiktiven Ingests aus sechs Services Tasks, drei User-Tasks, drei Send-Tasks, sieben Script-Tasks, einer Business Rule Task und zwei Gateways. Die Service-Tasks nutzen die RESTful-Web Services und innerhalb der Script-Tasks werden mittels weniger javascript-Anweisungen die Ergebnisdaten der RESTful-Web Services ausgewertet und ggf. deren Daten oder Metadaten als Variablen in den Prozess aufgenommen. Die Send-Tasks wurden für das Versenden der E-Mail genutzt. Die User-Tasks repräsentieren die Interaktionen des Nutzers mit dem Prozess über die webbasierte Benutzeroberfläche. Mit Gateways werden Entscheidungen modelliert, die den Prozessverlauf beeinflussen. In diesem Fall wurde geprüft, ob eine BagIt-URI eingegeben wurde und ob das SIP unverändert als API übernommen werden kann oder migriert werden muss. Die im AUdS-Vortrag enthaltenen Videos mit Untertiteln veranschaulichen die Umsetzung des fiktiven Ingest.

.....
27 <https://swagger.io> [Abruf 4. 10. 2019]

28 <https://camunda.com/download/modeler/> [Abruf 4. 10. 2019]

29 <https://camunda.com/de/> [Abruf 4. 10. 2019]

Die nachfolgende Abbildung zeigt die Konfiguration einer einen RESTful-Web Service aufrufenden Service Task. Der Script-Ausschnitt enthält einige javascript-Anweisungen zur Übernahme des Response (JSON Object) des RESTful-Web Services in die Variable „CheckMimeType“. Im Hintergrund ist ein Abbild des gesamten BPMN-Modells schemenhaft sichtbar.

Abb. 3: Seite 164

Fazit

Archivar und Informatiker können sich auf die Umsetzung der fachlichen und archivtechnischen Aufgaben konzentrieren, denn eine große Herausforderung besteht in der Modellierung und Implementierung von Workflows, die mit BPMN und einer BPM-Engine erheblich erleichtert wird. Die webbasierte Benutzeroberfläche der BPM-Engine und die mit APEX gestaltete und ebenfalls webbasierte Benutzeroberfläche ermöglichen die zeitnahe und zeitgemäße Umsetzung fachlicher Anforderungen ohne externe Unterstützung. Anspruchsvolle RESTful-Web Services können von einem erfahrenen Programmierer implementiert werden oder als klar formulierbarer Auftrag an einen Dienstleister vergeben werden.

Schlussendlich obliegen Archivar und Informatiker die Modellierung von Workflows und die Beschreibung von Funktionalitäten, die vorzugsweise mit RESTful-Web Services umgesetzt werden sollten. Dabei können Sie Ihre Anforderungen sukzessive entwickeln, denn zu Projektbeginn kann niemand die fachlichen Anforderungen bis ins letzte Detail ausformulieren.

Die ausschließliche Nutzung von RESTful-Web Services garantiert ein Höchstmaß an Flexibilität und gewährleistet eine bisher selten erreichte Produktunabhängigkeit.

References

- Reference Model for an Open Archival Information System (OAIS); <https://public.ccsds.org/pubs/650x0m2.pdf> [Abruf 4. 10. 2019]
- Object Management Group, Business Process Model and Notation (BPMN), <http://www.bpmn.org> [Abruf 4. 10. 2019]
- Representational state transfer (REST), https://en.wikipedia.org/wiki/Representational_state_transfer [Abruf 4. 10. 2019]
- software development (Dev) and information technology operations (Ops). <https://en.wikipedia.org/wiki/DevOps> [Abruf 4. 10. 2019]

The image shows a screenshot of the Camunda BPMN editor. On the left, a flow diagram is visible with several tasks: 'Start Prozess', 'CheckMimeType', 'Output Parameter', 'Script', and 'End Prozess'. Annotations in German explain parts of the diagram: 'Anzahl der Payload Data Notwendig für die Paralle...' points to a 'Parallel Gateway' (diamond); 'Bagit Payload ermitteln' points to the 'Script' task; 'JHove Identifikation durchführen' points to the 'Script' task's configuration; and 'Fremde Anweisung für den PayloadCode' points to the 'Script' task's configuration area.

The right side of the image shows the 'Properties Panel' for the 'Script' task. It contains the following sections:

- Details**
 - Connector Id: http-connector
- Input/Output**
 - Input Parameters: method: Text, headers: Map, url: Script
 - Output Parameters: CheckMimeType: Script
- Output Parameter**
 - Name: CheckMimeType
 - Type: Script
 - Script Format: javascript
 - Script Type: Inline Script
 - Script:

```

var json = S[connector.getVariable("response")];
var suffix = "Sub_" + connector.getVariable("loopCounter");
connector.setVariable("jHoveResult" + suffix, json);
connector.setVariable("mime-type" + suffix, json.prop("mime-type").value());
connector.setVariable("url" + suffix, json.prop("url").value());
json.prop("mime-type").value();
                    
```

Abb. 3: Service Task – 300dpi-ServiceTask.jpg

AUDS-CAMP
VOLKSZÄHLUNG 2021
UND DIE ROLLE
DER DIGITALEN ARCHIVE

ARCHIVCAMP

„VOLKSZÄHLUNG 2021 UND ROLLE DER DIGITALEN ARCHIVE“

Zusammenfassung

Christian Keitel

Im Plenum des Archivcamps „Volkszählung 2021 und Rolle der digitalen Archive“ zeigten die beiden Vorträge, dass sich die aktuellen Herausforderungen, denen sich die Archive in den Statistischen Ämtern gegenüberstehen, zwar erheblich unterscheiden. Zugleich sind die anzustrebenden Ziele und die dabei zu bewältigenden Schwierigkeiten sehr ähnlich. Die Teilnehmerinnen und Teilnehmer aus Deutschland und Tschechien waren sich einig, dass Statistiken eine sehr wichtige Rolle in der Überlieferungsbildung der zuständigen Archive spielen sollten. In den meisten Fällen dürfte sich diese Aufgabe vor allem den staatlichen Archiven stellen. Während es in den vergangenen Jahrzehnten gerade in Deutschland immer wieder auch andere Stimmen gab, die die Archivwürdigkeit von Statistiken mehr oder weniger direkt bestritten hatten, scheint dieses Thema nun erledigt zu sein. In der Tschechischen Republik stellt sich die Situation anders dar. Die Archivwürdigkeit der Statistik wurde nie in Frage gestellt (jeder Totalitarismus liebt personenbezogene Daten). Problematisch war aber die Menge, da die Erhebungsunterlagen einer Volkszählung Regalböden in der Länge von mindestens einem Kilometer belegen. Mit der Digitalisierung der Volkszählung ging im 21. Jahrhundert eine Lösung für den Umfang einher. Zugleich ergab sich ein neues Problem, nämlich der Schutz personenbezogener Daten und die Archivwürdigkeit anonymisierter Daten.

Ein noch umstrittener Bereich ist die Anbietungspflicht statistischer Unterlagen. In Deutschland schien sie nach Grundsatzentscheidungen des Bundesbeauftragten für Datenschutz und einigen Landesbeauftragten in den 1980er und 1990er Jahren festzustehen. In der alltäglichen Praxis haben die Statistischen Ämter aber in den letzten Jahren die Abstimmung erheblich erschwert, indem sie versuchten, die Aufbewahrungsfristen in den Ämtern sehr hoch zu setzen und damit eine Anbietung de facto erheblich hinauszuschieben. Zeitgleich bauen die Statistischen Ämter derzeit in Deutschland sogenannte Forschungsdatenzentren auf, die in immer größerem Umfang auch die Daten älterer Erhebungen zur Nutzung anbieten.

Geradezu legendär sind die Auseinandersetzungen zwischen Statistischen Ämtern und Archiven in Sachen Datenschutz. Karolína Šimůnková belegte in ihrem mit Martin Šisler gemeinsam erarbeiteten Vortrag sehr eindrücklich, dass gerade in osteuropäischen Ländern der Datenschutz eine außergewöhnlich hohe Priorität besitzt. In Deutschland muss das Grundrecht des Datenschutzes gegen das ebenfalls im Grundgesetz verankerte Grundrecht der Wissenschafts-

freiheit abgewogen werden. 100% Datenschutz könnte, so die Befürchtung, zu 100% Löschung der Unterlagen führen und damit in einem kompletten Gedächtnisverlust der Gesellschaft münden. Letztlich wurde dieses Spannungsverhältnis durch die Verhandlungen in den Parlamenten während der gemeinsamen Entwicklung der Archiv- und der Datenschutzgesetze aufgehoben.

Neben der grundsätzlichen Anbietungspflicht der statistischen Unterlagen wird in beiden Ländern auch diskutiert, ob nicht bestimmte Merkmale von der Anbietungspflicht ausgeschlossen werden sollten. In Deutschland wurde in den 1980er und 90er Jahren festgelegt, dass die die einzelnen Personen identifizierenden Hilfsmerkmale (Merkmale, die nur zur Erhebung selbst benötigt werden, aber selbst nicht statistischen ausgewertet werden) gelöscht werden müssen und dem Archiv nicht angeboten werden dürfen. Auch die tschechischen Kolleginnen und Kollegen konnten über vergleichbare, teilweise auch noch weitergehende Löschvorhaben berichten.

Einigkeit bestand unter den Teilnehmerinnen und Teilnehmern, dass Archive daher einen langen Atem brauchen, bis sie die ersten Statistiken in ihre Magazine holen können. Deutlich wurde aber auch, dass sie über ausreichend Engagement verfügen, um auf die lange Sicht erfolgreich sein zu können.

STATISTIK IM ARCHIV – EINE SCHWIERIGE BEZIEHUNG

Christian Keitel

Die Beziehungen zwischen den Statistischen Ämtern und den klassischen Archiven sind vielschichtig. Beide sind ein Ergebnis frühneuzeitlicher Regierungskunst. Jacob von Rammingen erklärte bereits im 16. Jahrhundert, dass alle Herrschaft auf der Erfassung ihrer Rechte (Renovatur) und der sicheren Verwahrung ihrer Akten in den Registraturen und Archiven aufbaue.¹ Die Renovaturen schlugen sich in den Urbaren nieder, die als früheste Formen der Landesbeschreibung verstanden werden können. Diese Aufgabe der Landesbeschreibung wurde dann zunehmend durch Statistiken umgesetzt. Mit anderen Worten sollten die Statistiken die Untertanen sowie die Städte, Dörfer und Gebiete beschreiben und dadurch beherrschbar machen, während die Archive die verschriftlichten Grundlagen der Herrschaft sicher zu verwahren hatten. Seit jeher stehen daher Statistische Ämter und Archive in einem engen Zusammenhang. Diese Beziehungen erhielten im 20. Jahrhundert eine neue Qualität. Nun fragte sich die Kollegen, ob und wie Statistiken auch in die Archive gelangen sollten.

Als Startpunkt für die Diskussion der archivischen Bewertung kann wenigstens in Deutschland ein Vortrag von Georg Hille gelten, den dieser auf dem 2. Deutschen Archivtag 1900 in Dresden gehalten hatte. Bereits in diesen ersten Überlegungen spielten Volkszählungsunterlagen eine Rolle. Hille war der Meinung, dass diese problemlos kassiert werden könnten: „Mit den von den Statistischen Aemtern ausgenutzten Volkszählungszetteln dürfen wir nach meiner Ansicht die Staatsarchive nicht belasten.“² Mit dem schönen, sonst unbekanntem Adjektiv „ausgenutzt“, benennt Hille sehr anschaulich seine Vermutung, dass sämtliche Informationen von diesen Volkszählungszetteln bereits extrahiert worden seien. In der Diskussion widersprach der Kollege Wolfram aus Metz. Die Volkswirtschaftslehre könne sehr wohl noch Interesse an diesen Zetteln haben. Da andererseits nicht alle Zettel erhalten werden könnten, plädierte er für den Erhalt einer Zählung pro Generation. Der wie Hille aus Schwerin stammende Wilhelm Grotefend sah einen

.....

1 „daß allen Herrschafften unnd Oberkeiten an einem fleissigen Renovieren irer Regalien, Güter und Rechten und dann an einer ordenlichen Registratur irer Brieff und Schrifftten sehr vil und hoch gelegen sein will, one welche zwey nicht statlich nützlich noch vollkommen inn den Regimenten derselben Cantzley und Raytcamer gehaußt werden mag noch kan.“ Jacob von Rammingen, Der rechten künstlichen Renovatur Eigentliche unnd gründtliche Beschreibung..., Heidelberg 1571 [S. 9 f.].

2 Georg Hille, Die Grundsätze der Aktenkassation, in: Korrespondenzblatt des Gesamtvereins der deutschen; Geschichts- und Alterthumsvereine 49 (1901), S. 26–31, hier S. 30.

Nutzen für die Familiengeschichte und forderte einen Erhalt aller Volkszählungsunterlagen.

Familienforschung und Volkswirtschaftslehre zählten um 1900 nicht zu den willkommenen Nutzergruppen, weshalb sich vielleicht nicht alle, sicher aber die meisten Archive darauf beschränkten, die tabellarischen Auswertungen, nicht aber die Erhebungsunterlagen zu erhalten. Bis in die 1950er Jahre hinein ist auch kein archivfachlicher Artikel bekannt, der sich eingehend mit der Bewertung von Statistiken beschäftigt hätte. Der Anstoß dazu musste von außen kommen. Die als Bibliothekarin beim Statistischen Bundesamt angestellte Lisa Kaiser machte im *Archivar* 1953 darauf aufmerksam, dass Statistiken sehr wohl archivwürdig sein könnten.³ Im selben Jahr befand der aus Württemberg stammende Walter Grube, dass Erhebungsbögen zu archivieren seien, Lochkarten allerdings nicht.⁴ Diese Zurückhaltung gegenüber den neuen Technologien manifestierte sich ein weiteres Mal in den frühen 1960er Jahren. Damals meinte das deutsche Bundesarchiv, die neu entstandenen Magnetbänder der Statistischen Ämter nicht selbst bewerten zu können. Lieber gab man diese Aufgabe an den Diplom-Soziologen Klaus Döll weiter. In einem Gutachten kam dieser denn auch prompt zum Ergebnis, dass Teile des Materials archivwürdig seien, was wiederum in weiten Kreisen der Archivwelt mit Kopfschütteln und Ablehnung quittiert wurde.⁵ Erst Rainer Stahlschmidt gelang es 1977 die Archivwürdigkeit von Statistiken so zu belegen, dass nun auch die Kolleginnen und Kollegen folgen konnten⁶. Seitdem gab es über die Bewertung von Statistiken in den deutschsprachigen Fachkreisen keinen grundsätzlichen Dissens. Großzählungen wie Volkszählungen oder Gebäude- und Wohnungszählungen gelten gemeinhin als archivwürdig. Auch andere Statistiken können archivwürdig sein. Bei vielen Statistiken dürften aber auch die früher auf Papier, nun in Online-Informationssystemen publizierten aggregierten Informationen genügen. 2008 konnte eine von den Leiterinnen und Leitern der Staatlichen Archive eingesetzte Arbeitsgruppe einen Bewertungskatalog der etwa 500 Bundesstatistiken vorlegen, der dann 2015 von einer weiteren Arbeitsgruppe leicht modifiziert bestätigt wurde. Beide Male waren etwa 70 Statistiken archivwürdig.⁷

Obwohl daher bereits in dem allerersten Fachbeitrag zur Bewertung die Archivwürdigkeit von Statistiken diskutiert worden war, dauerte es noch etliche Jahrzehnte, bis sich die Mehrheitsmeinung in der Fachgemeinschaft änderte. Da archivistische Bewertungsentscheidungen stets irreversibel oder in anderen

3 Lisa Kaiser, Zur Frage der Archivierung von Aktenbeständen staatlicher statistischer Ämter im Bundesgebiet, in: *Archivar* 6 (1953), Sp. 113-124.

4 Walter Grube, Das Problem der Massenakten, in: *Der Archivar* 7 (1954), Sp. 253-262.

5 Döll, Klaus, Die Aufbewahrung sozialwissenschaftlich wichtiger Massendaten, mschrftl. Köln o. J. [1965].

6 Rainer Stahlschmidt, Zur Archivierung des Datenmaterials der amtlichen Statistik in Nordrhein-Westfalen, Veröffentlichungen der staatlichen Archive des Landes Nordrhein-Westfalen Reihe E (Beiträge zur Archivpraxis) Heft 3, Siegburg 1980.

7 KLA-Arbeitsgruppe Bewertung von Statistikunterlagen (Hrsg.), Bewertung von Statistikunterlagen. Abschlussbericht Juni 2016, http://www.bundesarchiv.de/DE/Content/Downloads/KLA/abschlussbericht-bewertung-statistikunterlagen.pdf?__blob=publicationFile [Abruf 26. 4. 2019].

Worten sehr weitreichend sind, sollten wir uns gerade bei geänderten Einschätzungen in diesem Themenfeld fragen, weshalb sich die heutige Meinung nicht schon viel früher durchsetzen konnte. Bei den Statistiken lassen sich wenigstens drei Motive benennen:

- Geänderte Annahmen über Archivnutzer
- Geänderte Bewertungsmotive
- Geänderte Einstellung zur Technik.

Zu den Annahmen über Archivnutzer: In der Diskussion um den Hille-Vortrag wurde eine mögliche Archivwürdigkeit mit einer Nutzung durch Volkswirtschaftler oder Familienforscher begründet. Das Bundesarchiv sah in den 1960er Jahren vornehmlich eine Verwendung der Magnetbänder durch die Sozialwissenschaftler, weshalb man sich auch selbst die Bewertung nicht zutraute. Selbstkritisch kann angemerkt werden, dass die einseitige Ausrichtung auf professionelle Historiker dazu führte, dass Sozialwissenschaftler, Volkswirtschaftler und Familienforscher von den Archivaren nicht als potentielle Nutzer gesehen wurden, weshalb die Bewertungsentscheidungen entsprechend negativ ausfielen. Der geänderte Auftrag der Archive lässt sich heute gut an den Archivgesetzen ablesen, die seit der zweiten Hälfte der Achtziger Jahre in rascher Folge erlassen wurden. Nutzung wird nun zu einem grundlegenden Bürgerrecht. Historiker sind daher nur eine unter mehreren denkbaren Nutzergruppen, sie haben ihre spezielle Privilegierung weitgehend verloren.⁸

Zu den Bewertungsmotiven: Die führenden Archivwissenschaftler des letzten Jahrhunderts haben sich in fast allen Arbeiten vorwiegend mit records respektive Akten beschäftigt. Ob wir nun in die Werke von Hillary Jenkinson, Theodore Schellenberg, Adolf Brenneke, Heinrich Otto Meisner, Gerhard Enders oder Johannes Papritz schauen, stets ging es um die Frage, wie sich das Verwaltungshandeln in Akten und records niederschlug. Archive stehen dann vor der Aufgabe, Verwaltungshandeln in Auswahl zu dokumentieren und seine spätere Überprüfung zu ermöglichen. Statistiken dienen aber ganz anderen Zwecken. Sie versuchen seit der Frühen Neuzeit, Realität abzubilden, sie also überhaupt erst verwaltbar zu machen. Mit anderen Worten erfüllten Statistiken nicht die von den genannten Archivwissenschaftlern formulierte Anforderung an potentiell archivwürdige Unterlagen. Zugleich können wir in den Vereinigten Staaten von Amerika heute ein wesentlich differenzierteres Vorgehen beobachten. Das amerikanische Nationalarchiv hat die Ziele der Überlieferungsbildung prägnant in seiner Appraisal Policy formuliert.⁹ Archiviert werden sollen:

- Records documenting the rights of citizens
- Records documenting the actions of Federal officials
- Records documenting the national experience

Nur der als zweites genannte Punkt entspricht den Annahmen der erwähnten Archivwissenschaftler. Statistiken müssen dagegen zu den Records gezählt werden, die die nationalen Erfahrungen dokumentieren. In der deutschen Dis-

8 Vielleicht können die Regeln zur archivischen Sperrfristenverkürzung noch als Überreste der alten Privilegien erscheinen.

9 National Archives and Records Administration, Appraisal Policy of the National Archives and Records Administration, NARA 1441, December 6, 2005, https://www.fgdc.gov/organization/working-groups-subcommittees/hdwg/Appraisal/NARA_Appraisal/at_download/file [Abruf 29.5.2019].

kussion können in diesem Zusammenhang die Dokumentationsprofile genannt werden, in denen ebenfalls versucht wird, wesentliche Themenfelder unserer Gegenwart im Archiv abzubilden.¹⁰ Schließlich gibt es mit den Unterlagen, die die Rechte der Bürger dokumentieren, noch eine dritte Gruppe, die sich ebenfalls klar von den Verwaltungsunterlagen abhebt. Beim amerikanischen Nationalarchiv können hier die Datenbanken zu den gefallen Soldaten genannt werden, die die NARA ins Internet gestellt hat. Vermutlich können wir zu den Rechten der Bürger auch das Recht zählen, die eigene Geschichte rekonstruieren zu können. Jedenfalls wird ebendieses Recht derzeit sehr vehement von ehemaligen Heimkindern eingefordert. Gerade in diesem Punkt lässt sich dann wieder ein Bogen zu den Statistiken schlagen, denn die im britischen Nationalarchiv verwahrten Volkszählungslisten sind zentral für alle Fragen der Familienforschung.¹¹ Die Volkszählungen enthalten als wesentliche Identifikationsmerkmale die Namen der befragten Personen. Mit anderen Worten bieten sie eine sehr attraktive nicht-statistische Nutzung an. Auch bei den digitalen Unterlagen ist daher die von Theodore Schellenberg getroffene Unterscheidung von Primärwert und Sekundärwert unverzichtbar.¹²

Zum Technik-Problem: Seit jeher standen die Statistischen Ämter technischen Neuerungen sehr aufgeschlossen gegenüber. Erfunden haben sie die Lochkartentechnik zwar nicht. Seit 1805 wurden Webstühle und auch andere Maschinen mit Lochkarten gesteuert. 1890 wurden jedoch weltweit erstmals Lochkarten zu Informationssammlung und -verarbeitung eingesetzt – bei der Amerikanischen Volkszählung. 1910 wurden diese Maschinen erstmals in Deutschland eingesetzt – bei der Verarbeitung der Volkszählung in den statistischen Ämtern.¹³ 1961 setzten die Statistischen Ämter erstmals Computer zur Auswertung der Lochkarten ein und 1970 wurden erstmals die Zählunterlagen durch einen Computer eingelesen. Statistische Ämter waren daher beim Einsatz neuer Technologien häufig Pioniere innerhalb ihrer Verwaltungen.

Archive haben sich dagegen über viele Jahre eine vornehme Zurückhaltung auferlegt, sobald es um Lochkarten und Computer ging. Einerseits hat man sich mit diesen Maschinen bereits in den 1970er Jahren eingehend beschäftigt. Andererseits sollte es bis in unser Jahrhundert dauern, bis Archive in größerem Umfang in die digitale Archivierung einstiegen und sich dadurch auch intensiv mit der modernen Technik beschäftigten. Mitte der 1970er Jahre hat die damalige Landesarchivdirektion Baden-Württemberg für die Lochkarten des Statistischen Landesamts eine pauschale Vernichtungsgenehmigung ausgestellt, da mit der

10 Vgl. Christian Keitel, Eine andere Art der Dokumentation. Anmerkungen zur Bewertung umfassender Informationssysteme, Stuttgart 2011, http://www.landesarchiv-bw.de/sixcms/media.php/120/52529/Workshop_Keitel_andere_Art.pdf [Abruf 26.4.2019].

11 <http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/census-records/> [Abruf 9.4.2019].

12 Theodore R. Schellenberg, Modern archives. Principles and techniques, Chicago 1956, S. 16.

13 Zum Folgenden vgl. Christian Keitel, Baden-württembergische Archivverwaltung beginnt mit der elektronischen Archivierung. Volkszählung 1970 als erstes digitales Archivale im Staatsarchiv Ludwigsburg archiviert. In: Der Archivar 57 (2004) und ders., 35 Jahre „Elektronische Archivierung“ in der Archivverwaltung Baden-Württembergs: Ein Lernprozess, in: Archive in Bayern Bd. 6 (2010), S. 11–32.

damaligen Technologie diese Karten nicht nachnutzbar waren. Heute wären wir dazu in der Lage. Wir könnten jede Lochkarte in einen Datensatz überführen. Leider müssen wir daher heute einen Totalverlust maschinenlesbarer Überlieferung für ein halbes Jahrhundert, also die Zeit von 1910 bis 1961 konstatieren.

Nicht viel anders sah es zunächst in den Vereinigten Staaten aus. 1936 waren sich die Vertreter von Nationalarchiv und Statistikbehörde darin einig, dass die Lochkarten erschöpfend ausgewertet wären und daher nicht archiviert werden müssten.¹⁴ Bevor das Nationalarchiv 1970 als weltweit erstes Archiv digitale Unterlagen übernehmen konnte, mussten die Kollegen Meyer Fishbein und Everett Alldredge die Archivwürdigkeit maschinenlesbarer Unterlagen erst nachweisen.¹⁵

Die lange andauernde Zurückhaltung gegenüber diesen neuen Formen der Informationsauszeichnung kann sicher zu großen Teilen auf die fremdartigen Maschinen zurückgeführt werden, ob sie nun Lochkartenmaschinen oder Computer hießen. Aber auch die von diesen Maschinen produzierten Ergebnisse unterscheiden sich doch erheblich von allem, mit dem die Archive bislang zu tun hatten.

Zunächst liegen uns Statistiken in aller Regel in sehr unterschiedlichen Formen vor. Die originalen Erhebungsunterlagen wurden bislang noch stets auf Papier verteilt und wieder eingesammelt. Diese Informationen werden seit 1970 automatisch von einem Computer eingelesen. Es entstehen dann Rohdaten, die inhaltlich weitgehend den Papierbögen entsprechen. Dennoch finden die Statistiker noch zahlreiche Unstimmigkeiten, die im Zuge der Plausibilisierung beseitigt werden sollen. Im Ergebnis entstehen plausibilisierte Roh- oder Mikrodaten. Diese werden gerne aggregiert, also auf der Ebene einzelner zumeist geographischer Einheiten wie Orte, Kreise oder Bundesländer feldweise zusammengefasst. Die Informationen liegen also mehrfach vor. Zugleich unterscheiden sich diese Ausgaben deutlich.

In dieser Situation sollten wir uns fragen, ob die Archive eher das Ausgangsmaterial auf Papier oder die digitalen Exzerpte übernehmen sollten. Das Landesarchiv Baden-Württemberg konnte 1973 2.300 Meter papierne Erhebungsunterlagen übernehmen. Bekanntlich folgten die Archive in den der Informationsverarbeitung vorangehenden Zeiten der Prämisse der Datenträgerarchivierung. Originale wurden und werden dabei stets den möglichen Kopien vorgezogen. In unserem Fall waren die papiernen Erhebungsunterlagen die Originale. Waren also die davon abgeleiteten Daten nur billige Kopien? Walter Grube hatte sich wie erwähnt 1954 für das Papier und gegen die Lochkarten ausgesprochen. Nun liegen die Daten aber bereits digital vor. 2002 konnten wir alle diese Informationen auf 3 DVDs übernehmen. Es ist offenkundig, dass das Papier in diesem Fall wesentlich höhere Lagerungskosten verursacht als die digitalen Daten. Hinzu kommt, dass sich diese fast mühelos digital auswerten lassen, während vor einer Auswertung der Erhebungsbögen äußerst umfangreiche Digitalisierungsvorhaben durchgeführt werden müssten. Aus Sicht der Nutzer sprach daher alles für eine Archivierung der digitalen Daten. Hinzu

.....
14 Meyer Fishbein, *Appraising Information in Machine Language Form*, in: *The American Archivist* 35 (1972), S. 35–43, hier S. 36.

15 Everett Alldredge, *Inventoring Magnetic-Media Records*, in: *The American Archivist* 35 (1972), S. 337–345.

kamen die deutlich niedrigeren Kosten. Die baden-württembergische Archivverwaltung entschied sich daher für den Erhalt der digital gespeicherten Information – und damit in diesem konkreten Fall gegen den Erhalt des Papiers. Für manche Kolleginnen und Kollegen bedurfte es Anfang des letzten Jahrzehnts einige Überwindung, um der weitgehenden Vernichtung der papiernen Erhebungsbögen (mit Ausnahme einiger Belegexemplare) zugunsten der digitalen Daten zuzustimmen.

Dabei war dies nicht die einzige Zumutung. Nun stellte sich nämlich die Frage, ob die Landesarchivverwaltung eher die unplausibilisierten Rohdaten oder die plausibilisierten Rohdaten übernehmen sollte. Die unplausibilisierten Rohdaten bilden unmittelbar die Antworten der befragten Einwohner ab. Die plausibilisierten bilden dagegen nach Meinung des Statistischen Amtes eher die damaligen Realitäten ab. Was war nun vorzuziehen? Erneut erscheint ein Zwiespalt. Die klassische archivarische Reaktion dürfte zum Erhalt der nicht-plausibilisierten Rohdaten tendieren. Niklas Bütikofer hat aber überzeugend dargelegt, dass eine Plausibilisierung sehr weitgehend von der Kenntnis heutiger Rahmenbedingungen abhängt, also von etwas, das in wenigen Jahren nicht mehr greifbar sein dürfte.¹⁶ Mit anderen Worten können Plausibilisierungen in einigen Jahren nicht mehr nachgeholt werden. Wenn wir versuchen, die Perspektive der künftigen Nutzer einzunehmen, liegt die Antwort nahe: Statistische Unterlagen werden dazu verwendet, vergangene Realitäten abzubilden, nicht aber, um fehlerhafte Antworten lesen und vielleicht mangels Kontextinformation gar nicht erkennen zu können. Letztlich haben wir uns daher für die Übernahme der plausibilisierten Unterlagen entschieden. Zugleich haben wir uns aber erneut gegen die Information entschieden, die wir bei klassischer Sichtweise als Original bezeichnen könnten. Sowohl bei der Frage „Erhebungsunterlagen oder digitale Derivate“ als auch bei der Frage der Plausibilisierung fiel die Entscheidung daher anders aus, als sie bei „normalen“ Akten getroffen worden wäre. Solche Abweichungen von der eingeübten Regel erfordern Überwindung und in dieser Überwindung dürfte ein wesentlicher Grund für die relativ späte archivarische Beschäftigung mit Statistiken liegen.

Gegenüber den Erhebungsunterlagen auf Papier und den nicht plausibilisierten Rohdaten kann eine bevorzugte Archivierung der plausibilisierten Rohdaten gut begründet werden. Dasselbe gilt aber auch für die Abgrenzung der plausibilisierten Rohdaten von den aggregierten Formen. Jede Aggregation zerstört die Zusammenhänge zwischen den Teilantworten einer befragten Person. Nach erfolgter Aggregation können wir nicht mehr sagen, dass es in Musterdorf 1970 einen siebzehnjährigen Schornsteinfegerlehrling gegeben haben muss, der über 20 Kilometer zu seinem Ausbildungsbetrieb pendeln musste. Wir wissen nur noch, dass es x 17jährige, y Schornsteinfegerlehrlinge und z Menschen gab, die über 20 Kilometer zu ihrem Ausbildungsbetrieb pendelten. Eine auf die Ergebnisse mehrerer Fragen bezogene Abfrage, im technischen Sinne also eine Felder übergreifende Abfrage, ist also nicht mehr möglich und auch nicht mehr rekonstruierbar. Dagegen sind umgekehrt Aggregationen aus den Rohdaten stets möglich. Erneut gilt: Mit Blick auf die künftige Nutzung kann eine Entscheidung zugunsten der Rohdaten begründet werden. Sobald wir eine künftige

.....
 16Niklaus Bütikofer, Archivierung von statistischen Daten, in: Studien und Quellen 16/17 (1990/91), Bern 1991, S. 263–292, hier S. 275.

Nutzung mit übergreifenden Fragestellungen vermuten, sollten wir die Rohdaten erhalten. In anderen Fällen genügen sicher auch die aggregierten Daten.

Am Anfang dieses Beitrags wurden Archive und Statistische Ämter als Geschwister beschrieben, die, wie das bei Geschwistern eben häufig vorkommt, sich zwar ähnlich, zugleich aber oft nicht einer Meinung sind. Nicht nur die Archive, sondern auch die Statistischen Ämter haben in den letzten Jahrzehnten immer wieder gezeigt, wie fremd ihnen ihr Gegenüber ist. Nachdem sich infolge von Rainer Stahlschmidts bahnbrechender Untersuchung seit den späten 1970er Jahren die Archive der Statistik zuwandten, wurden sie nicht unbedingt mit offenen Armen empfangen. Zunächst stand die Zulässigkeit von Volkszählungen überhaupt in Frage. In einem Grundsatzurteil begründete das Bundesverfassungsgericht in Deutschland den Datenschutz und setzte den Behörden in ihrem Handeln enge Grenzen. Konkret bedeutete dies, dass in der zweiten Hälfte der 1980er Jahre sowohl Statistikgesetze als auch Archivgesetze diskutiert und verabschiedet werden mussten. Eigentlich sollte man meinen, dass nun die Anbietungspflicht aller Behördenunterlagen definitiv festgestellt worden sei. Als allerdings Anfang der 1990er Jahre das Bundesarchiv damit beginnen wollte, Statistiken von Statistischen Bundesamt zu übernehmen, wurde von statistischer Seite diese Anbietungspflicht mit Verweis auf den Datenschutz bestritten. Erst ein Schiedsspruch des Bundesdatenschutzbeauftragten setzte die grundsätzliche Anbietungspflicht durch. In Baden-Württemberg wiederholte sich derselbe Prozess einige Jahre später auf Landesebene.¹⁷

Diese Prozesse führten für die Archive zu einer doppelten Klarstellung: Einmal waren nun statistische Unterlagen grundsätzlich anbietungspflichtig. Allerdings war ein kleiner Teil dieser Unterlagen von der Anbietung ausgenommen. Die sogenannten Hilfsmerkmale, also Angaben, die zwar nicht selbst zur Statistik respektive Zählung gehörten, die aber für deren Durchführung benötigt wurden, wie zum Beispiel Vor- und Nachnamen, Hausnummern und Zählbezirke, diese Hilfsmerkmale also mussten auf Anweisung des Verfassungsgerichts gleich nach Datenerhebung noch im Statistischen Amt gelöscht werden. Sie konnten und durften den Archiven nicht angeboten werden.

Anfang des letzten Jahrzehnts stellte dann ein im Auftrag der Statistischen Ämter angefertigtes Gutachten fest, dass Statistiken in den Statistischen Ämtern aus rechtlichen Gründen wahrscheinlich nie umfassend genutzt werden können. Die Statistischen Ämter schlossen nun mit den Staatlichen Archivverwaltungen einen Rahmenvertrag ab, der die kontinuierliche Anbietung älterer Statistiken an die zuständigen staatlichen Archive regeln sollte. Dieser Rahmenvertrag hat übrigens auch die Einrichtung der bereits erwähnten Bewertungsarbeitsgruppen auf archivischer Seite nach sich gezogen. Auf archivischer Seite war daher das Feld bestellt. In der Zwischenzeit hatten die statistischen Ämter jedoch eine Reihe von Forschungsdatenzentren aufgebaut, die eine statistische Nutzung älterer Statistiken ermöglichten. Auf einmal war das statistische Interesse an diesen Daten sprunghaft angestiegen. Als dann 2014 von den Statistischen Ämtern eine zentrale Nachweisdatenbank für alle neuern Statis-

17 Wolf Buchmann, Michael Wettengel, Auslegung des Bundesstatistikgesetzes bei der Archivierung von Statistikunterlagen, in: *Der Archivar* 49 (1996), H. 1, S. 67-74; Udo Schäfer, Die Pflicht zur Anbietung und Übergabe von Unterlagen in der archivischen Praxis, in: Robert Kretschmar (Hg.), *Historische Überlieferung aus Verwaltungsunterlagen*, Stuttgart 1997, S. 40-46.

tiken aufgebaut wurde, setzten sie deren Aufbewahrungsfrist pauschal auf die maximal möglichen 30 Jahre hoch. Diese Frist sollte erst mit dem Nachweis in diesem System zu laufen anfangen. Eine 2005 erstellte Statistik könnte folglich frühestens 2044 angeboten werden (2014 + 30). In der aktuellen Diskussion stimmen die staatlichen Archivverwaltungen eine Reaktion auf diese aus ihrer Sicht nicht gerechtfertigte Vorgehensweise ab. Sie entzieht den interessierten Nutzern diese Unterlagen wenigstens teilweise. Die Statistiken können nicht im Kontext anderer zeitgenössischer Unterlagen ausgewertet werden. De facto wird ein Parallelarchiv aufgebaut, das von der gemeinsamen Bundes- bzw. Landesverwaltung zusätzlich finanziert werden muss. Schließlich besteht keine Gewähr, dass die so langen zurückgehaltenen digitalen Daten nach Ablauf dieser langen Frist überhaupt noch lesbar sind.

Vergleichbare Behördenarchive zeichnen sich auch im Gebiet der Vermessungsverwaltungen ab. In der Schweiz hat sich swisstopo sogar per Gesetz festschreiben lassen, manche Unterlagen erst nach 275 Jahren anbieten zu müssen.¹⁸ In Deutschland gibt es auch vergleichbare, bislang zum Glück weniger ambitionierte Ansätze. Auch hier ist eine Erhöhung der staatlichen Aufwände für digitale Erhaltung einerseits und eine Verschlechterung der Möglichkeiten für Nutzer andererseits zu erwarten.

Wo stehen also heute Archive und Statistische Ämter in Deutschland? Trotz aller Schwierigkeiten haben sich die beiden Seiten erheblich aufeinander zubewegt. Zwar hat es einige Jahrzehnte gedauert, bis sie den heutigen Zustand erreicht haben. Die zurückgelegten Distanzen sollten uns aber hoffen lassen, dass auch die verbleibenden Meinungsverschiedenheiten noch gelöst werden können.

.....
 18 Bewertungsentscheid Geo(basis)daten des Bundes (Projekt Ellipse, AAP), [https://www.bar.admin.ch/dam/bar/de/dokumente/bewertungsentscheide/Geobasisdaten%20Bewertungsentscheid%202016.pdf.download.pdf/Bewertungsentscheid%20Geo\(basis\)daten%20des%20Bundes%20\(Pjekt%20Ellipse,%20AAP\),%202016-02-19.pdf](https://www.bar.admin.ch/dam/bar/de/dokumente/bewertungsentscheide/Geobasisdaten%20Bewertungsentscheid%202016.pdf.download.pdf/Bewertungsentscheid%20Geo(basis)daten%20des%20Bundes%20(Pjekt%20Ellipse,%20AAP),%202016-02-19.pdf) [Abruf 29. 5. 2019].

ZUR PROBLEMATIK DER ELEKTRONISCHEN ARCHIVIERUNG VON VOLKSZÄHLUNGEN 2011 UND 2021 IN DER TSCHECHISCHEN REPUBLIK

Karolína Šimůnková, Martin Šisler

In den öffentlichen Archiven der Tschechischen Republik finden wir zur Zeit ca. 45 Bestände aus der Zeit 1869 – 2011 mit der Dominanz der statistischen Personaldaten. Im Nationalarchiv in Prag befinden sich mehrere Archivbestände, die die statistischen Daten enthalten. Weil das Nationalarchiv das Zentralarchiv für gesamtstaatliche Agenden ist, befinden sich gerade hier die grössten Bestände dieser Art. Seit Anfang des 20. Jahrhundert sind das mindestens sieben Archivbestände, sechs aus dem gesamtstaatlichen Zensus, dazu ein Zensus aus dem Jahre 1939 aus dem überwiegend von Deutschen bewohnten Grenzgebiet, das nach dem Münchener Abkommen an das Deutsche Reich abgetreten worden war. Aus diesen insgesamt sieben Bestände sind vier in Völligkeit (das bedeutet mit vollen unanonymisierten Personaldaten und für das ganze Gebiet des Staates¹) erhalten. Es sind Zensus 1930, 1950, 1970, 1991, das bedeutet jeder zweiter im 20jährigen Rhythmus. Es handelt sich um ziemlich grosse Bestände, Zensus 1950 bildet z.B. 3.042,5 Lfm. Zur Vollständigkeit ist es notwendig zu sagen, dass auf Grund des enormen Umfangs der Zensus aus dem Jahre 1961, der vollständig im Nationalarchiv erhalten war, im Jahre 1981 deutlich reduziert wurde, die Fragebogen wurden, mit der Ausnahme von 7 Bezirken (ca. 87 Lfm), vernichtet. Zensus 1961 gehört aber nicht zu der Reihe der Zensus im 20jährigen Rhythmus.

Den Entwurf des sachlichen Vorhabens des Gesetzes über die Bevölkerungs-, Gebäude- und Wohnungszählung im Jahre 2011 hat das Statistikamt zum ersten Mal im September 2006 ins interministerielle Anmerkungsverfahren vorgelegt. Der Entwurf wurde in zwei Varianten vorgelegt, die sich durch die Weise der Durchführung der Zählung unterscheiden. Mit Rücksicht auf die Kompliziertheit der Auseinandersetzung des Anmerkungsverfahrens bei den Entwürfen der Variantenlösung beantragte das Statistikamt im November 2006 die Aufhebung der Aufgabe und im März 2007 legte das Statistikamt ins neue Anmerkungsverfahren einen Entwurf des sachlichen Vorhabens ohne Varianten vor, der mit niedrigeren Haushaltsausgaben rechnete. Dieser Entwurf wurde

.....
1 Fragebogen und Statistiken, die sich auf das Gebiet der heutigen Slowakischen Republik beziehen, befinden sich in den slowakischen Archiven.

schon im September 2007 von der Regierung bewilligt und der Vorsitzende des Statistikamts wurde beauftragt den Gesetzesentwurf vorzulegen. Im März 2008 wurde der Entwurf ins interministerielle Anmerkungsverfahren vorgelegt und im September 2008 durch die Regierung verabschiedet. Das Zensusgesetz 2011 trat am 22. Juli 2009 in Kraft.

Wie schon oben erwähnt, wurden die Volkszählungslisten im Nationalarchiv nach einer zwanzigjährigen Periode erhalten. Aus dem Zensus im Jahre 2001 wurden nur die veröffentlichten Ergebnisse des Zensus in der Form der gedruckten Publikationen in das Nationalarchiv übernommen. Im Zensusgesetz 2001 wurde festgesetzt, dass die Fragebögen nach der Beendigung der Bearbeitung der Ergebnisse des Zensus vernichtet werden. Schon damals hat die Archivgemeinde dagegen bei der Vorbereitung des Gesetzes ohne Erfolg protestiert. Die zwanzigjährige Periode für die Übergabe ins Archiv sollte im Jahre 2011 erfüllt werden. Im April 2006 hatten das Nationalarchiv und die Abteilung der Archivverwaltung des Innenministeriums begründete Befürchtungen, dass nach der Beendigung des Zensus im Jahre 2011 genauso wie nach dem Zensus im Jahre 2001 verfahren wird. Bei der Vorbereitung des neuen Gesetzes haben sie sich darum dafür eingesetzt, dass im Gesetz statt der Bestimmung über die Vernichtung der Fragebögen angegeben wurde, dass die ausgefüllten Fragebögen nach der Bearbeitung der Ergebnisse des Zensus nach dem Archivgesetz behandelt werden. Das heißt, dass nach der Beendigung der Verarbeitung der Ergebnisse des Zensus eine Auswahl von Archivalien im Vernichtungsverfahren durchgeführt wird und die Archivalien im Nationalarchiv deponiert werden. Trotz aller Bemühungen waren die Anforderungen der Archivare nicht angenommen.

Der Grund der Ablehnung war insbesondere die Tatsache, dass eine der Schlüsselprioritäten des ganzen Zensus der Schutz der Personaldaten wurde. Die beim Zensus erworbenen Daten sollten also ausschließlich für statistische Zwecke ausgenutzt werden.² Die Fragebögen in der Druckausstattung sollten vernichtet werden, denn sie haben durch die Verarbeitung der zur Bearbeitung geleisteten Daten ihren Zweck erfüllt und es besteht kein rechtlicher Grund zu ihrer weiteren Erhaltung. Die bei der Verarbeitung in die elektronische Form übertragenen Fragebögen und die von den verpflichteten Personen erfüllten elektronischen Fragebögen sollten anonymisiert werden und spätestens bis nach drei Jahren seit dem entschiedenen Zeitpunkt zur dauerhaften Erhaltung im Nationalarchiv übergeben werden.

Im Juli 2011 stellte das Statistikamt das Material zum System der Archivierung der Fragebogen für das Nationalarchiv vor, in dem beschrieben wurde, welche Dokumente zur Archivierung übergeben werden, in welcher Weise sie technologisch vorbereitet und übergeben werden und wie die in den Fragebögen angeführten Personaldaten anonymisiert werden. Das Grundprinzip der Ar-

.....

2 Die allgemeine Rechtsquelle und die Beschränkung des statistischen Zwecks finden wir im Art. 20 der Verordnung (EG) Nr. 223/2009 des Europäischen Parlaments und des Rates vom 11. März 2009 über europäische Statistiken und zur Aufhebung der Verordnung (EG, Euratom) Nr. 1101/2008 des Europäischen Parlaments und des Rates über die Übermittlung von unter die Geheimhaltungspflicht fallenden Informationen an das Statistische Amt der Europäischen Gemeinschaften, der Verordnung (EG) Nr. 322/97 des Rates über die Gemeinschaftsstatistiken und des Beschlusses 89/382/EWG, Euratom des Rates zur Einsetzung eines Ausschusses für das Statistische Programm der Europäischen Gemeinschaften

chivierung der Fragebögen für die Übergabe an das Nationalarchiv war nach dem Material der Schutz der personenbezogenen Daten des Befragten. Die Anonymisierung der Daten in den Fragebögen sollte so durchgeführt werden, dass die entsprechenden Zeilen, in den die Daten vom Befragten geschrieben wurden, mit schwarzer Farbe bedeckt werden. Durch die Bedeckung mit schwarzer Farbe verstand man der Ersatz der Bitmap des entsprechenden Ausfüllungsrahmens durch eine Bitmap nur mit schwarzem Pixel und zwar in derselben Schicht, wo die Bitmap des Bildes des Fragebogens ist. Das Material rechnete mit der Anonymisierung der Identifikationsangaben und weiterer Angaben, die direkt eine persönliche Identifikation ermöglichen, wie Wohnungs- und Hausnummer, Straße, Zählbezirk, Nachname, Vorname, Geburtsnummer, Geburtstag und Geburtsmonat.

Seit Juni 2011 bis zum Januar 2012 hat das Personaldatenschutzamt die Kontrollen beim Statistikamt durchgeführt, die unter anderem auch auf die Weise der Anonymisierung der Fragebögen gezielt wurden, die ins Nationalarchiv übergeben werden sollen. Im Verlauf der Kontrolle ersuchte das Statistikamt die Kontrollgruppe des Personaldatenschutzamtes um Stellungnahme, welche Angaben nach den Vorstellungen des Personaldatenschutzamtes anonymisiert werden müssen. Im September 2011 bearbeitete das Personaldatenschutzamt die Stellungnahme, in der es verlangte, den ursprünglichen Umfang der anonymisierten Angaben wesentlich zu erweitern, wie ihn der Entwurf des Statistikamtes voraussetzte. Über den Rahmen dieses Entwurfs sollten solche Angaben, wie Gemeinde, Gemeindeteil, Postleitzahl, eingetragene Partnerschaft, Nationalität, Religionsgesellschaft und Beruf, anonymisiert werden. Bei der Beurteilung der Anonymisierung ging das Personaldatenschutzamt vom Gesichtspunkt aus, dass es nicht zulässig ist, mit Hilfe einer Kombination der im Nationalarchiv deponierten Angaben mit anderen, wenn auch öffentlich zugänglichen Angaben festzustellen, um welche Person es sich handelt. Diese Gefahr droht zum Beispiel bei Gemeinden mit weniger als 200 Einwohnern. Ein auf einer möglichen Identifikation von konkreten Angaben beruhender Missbrauch drohte nach dem Personaldatenschutzamt auch im Falle des Berufs in der Kombination mit der beruflichen Stellung. Im Falle von kleinen Religionsgesellschaften oder Religionsgruppen konnte diese in Verbindung mit anderen Angaben die Quelle einer möglichen Identifikation sein.

Danach wurde auf Veranlassung des Statistikamtes im Oktober 2011 eine Arbeitsverhandlung der Vertreter des Nationalarchivs, des Innenministeriums, des Personaldatenschutzamtes, des Statistikamtes und der Firma Hewlett-Packard einberufen, welche für die Durchführung der Anonymisierung sorgen sollte. Gegenstand der Verhandlung war die neue Anforderung des Personaldatenschutzamtes an eine grundsätzliche Erweiterung der Anonymisierung um weitere Angaben. Das Nationalarchiv machte die Anwesenden mit den Arten der Ausnutzung der Zählungsformulare aus den älteren Zensus bekannt und betonte insbesondere die umfangreichen Einsichtsagenden für die Bedürfnisse nicht nur der Staatsverwaltung, sondern auch der privaten Subjekte. Zu den erweiterten Anforderungen des Personaldatenschutzamtes haben die Vertreter des Nationalarchivs angegeben, dass die Angaben, wenn sie in solchem Umfang anonymisiert werden sollten, keinen Informationswert haben werden und das Nationalarchiv diese zur dauerhaften Erhaltung nicht übernimmt. Dazu haben sie noch darauf aufmerksam gemacht, dass jede beliebige Anonymisierung

der Angaben schon auf der Ebene des Urhebers, d. h. vor der Übergabe zur dauerhaften Deponierung ins öffentliche Archiv, im Widerspruch zum im Archivgesetz geregelten Verfahren ist. Das Nationalarchiv sah ein Problem auch in der Tatsache, dass einzelne Kopien im TIFF-Format übergeben werden sollten, in welchem es nicht möglich wäre, die Angaben auszusuchen. Das Personaldatenschutzamt hat bemerkt, dass das Nationalarchiv mit den Argumenten spät kommt und diese ins Gesetz über die Volkszählung projiziert werden sollten. In der gegenwärtigen Situation muss das Statistikamt nach dem Gesetz über das Zensus die Fragebögen zuerst anonymisieren und diese erst danach zur Deponierung ins Nationalarchiv übergeben. Das Innenministerium unterstützte die Stellungnahme des Nationalarchivs. Die Diskussion wurde mit dem Vorschlag abgeschlossen, die ganze Angelegenheit im Rahmen einer zweiseitigen Verhandlung zwischen dem Nationalarchiv und dem Personaldatenschutzamt abzuschließen, deren Ergebnis dem Statistikamt bekannt gegeben wird. Im Falle, dass beide Seiten zu keinem Kompromiss kommen und der Umfang der Anonymisierung im durch das Personaldatenschutzamt vorgeschlagenen Umfang bleibt, bietet das Statistikamt dem Nationalarchiv Dokumente zur Bewertung an und dieses wird entscheiden, ob sie archivwürdig sind, oder nicht.

Die nachfolgenden Verhandlungen zwischen dem Nationalarchiv und dem Personaldatenschutzamt wurden erfolglos beendet. Das Nationalarchiv argumentierte mit der Bestimmung des Archivgesetzes, die den Archiven ermöglicht, eine Anonymisierung der in den statistischen Dateien erhaltenen Personalangaben vor der Bereitstellung für die Öffentlichkeit durchzuführen. Die Anonymisierung bleibt so in der Verantwortung und in den Händen des zuständigen Archivs. Das Nationalarchiv schlug eine Kompromisslösung vor, und zwar, die Anonymisierung in zwei Phasen durchzuführen – in der ersten Phase würde das Statistikamt die Daten im Umfang seines ursprünglichen Entwurfs anonymisieren und in der zweiten Phase würde das Nationalarchiv die Anonymisierung vor der Bereitstellung für die

Öffentlichkeit durchführen. Schon durch die Annahme der Tatsache, dass die Anonymisierung nach dem Zensusgesetz 2011 noch vor der Übergabe der Fragebögen zur dauerhaften Erhaltung durchgeführt wird, macht das Nationalarchiv ein großes Zugeständnis, das im Widerspruch nicht nur zum Archivgesetz, sondern auch zur gegenwärtigen Praxis steht. In so anonymisierten Fragebögen wird es selbst für die Bedürfnisse der Staatsverwaltung nicht möglich sein, die Ermittlungsagenda durchzuführen.

Das Personaldatenschutzamt sah die einzige mögliche Lösung in der Novellierung des Zensusgesetzes, was aber nicht nur aus zeitlichen Gründen, sondern auch aus Prozessgründen in der gegebenen Zeit nicht mehr durchzuführen war, insbesondere in der Situation, wann das Nationalarchiv keine Anmerkungsstelle des Legislativprozesses war. Nach dem Personaldatenschutzamt dürfen die beim Zensus erworbenen Angaben in keinem Falle für Bedürfnisse der Staatsverwaltung dienen, die zu diesem Zweck andere Informationsquellen hat. Den Einwohnern, die verpflichtet waren, solche Informationen über sich anzugeben, welche die Staatsverwaltung nie zur Verfügung hatte, wurde versichert, dass diese Informationen nach drei Jahren vernichtet werden. Das Personaldatenschutzamt befürchtete auch weiterhin einen Missbrauch der erworbenen Daten zu Ungunsten konkreter einzelner Einwohner und hielt darum hinsichtlich des Umfangs der Anonymisierung an seiner ursprünglichen Stellungnahme fest. In

dieser Situation waren die Verhandlungen im November 2011 beendet, ohne zur gegenseitigen Vereinbarung zu kommen. Das Nationalarchiv gab dem Statistikamt bekannt, dass es leider gezwungen sein wird, auf dauerhafte Erhaltung der Dokumente, die keinen bleibenden historischen Wert haben, zu verzichten.

Die ganze Angelegenheit wurde aus dem Grund des sich nähernden durch das Zensusgesetz bestimmten Termins für die Übergabe der Daten zur Erhaltung ins Nationalarchiv im Herbst 2013 wieder eröffnet. Im September 2013 wurde darum auf Veranlassung des Statistikamts eine dreiseitige Verhandlung zwischen dem Statistikamt, dem Nationalarchiv und dem Innenministerium, diesmal schon ohne Teilnahme des Personaldatenschutzamts, verwirklicht. Als Grundlage der Verhandlung diente das vom Statistikamt ausgearbeitete Dokument mit dem Titel „Der vorgeschlagene Umfang der Anonymisierung der Daten aus dem Zensus 2011 für die Übergabe an das Nationalarchiv“, in dem das Statistikamt ersichtlich machte, welche Angaben es vorschlägt völlig zu anonymisieren, welche Angaben lesbar bleiben würden und welche nur teilweise anonymisiert würden. Im Vergleich mit dem ursprünglichen Entwurf des Statistikamts aus dem Jahre 2011 schlug das Statistikamt jetzt vor, die Anonymisierung um die Angaben Gemeindeteil und Postleitzahl zu erweitern. Die Stellungnahme des Personaldatenschutzamts aus dem Jahre 2011 wurde diesmal nicht in Erwägung gezogen, denn das Gesetz erlegt nach der Äußerung des stellvertretenden Vorsitzenden des Statistikamts diese Pflicht nicht auf. Das Statistikamt erwies sich bereitwillig, mit dem Nationalarchiv weiter über den Umfang der Anonymisierung der ausgewählten Datenfelder zu verhandeln und ersuchte das Nationalarchiv, schriftlich mitzuteilen, welche Datenfelder geeignet wären, bei der Erhaltung des Personaldatenschutzes, lesbar gelassen zu bleiben. Das Nationalarchiv konstatierte in schriftlicher Antwort mit Befriedigung, dass der aktuelle Entwurf des Statistikamts, den Schlussfolgerungen aus den vorherigen Verhandlungen gegenüber, eine bedeutsame Verringerung der anonymisierten Angaben bedeutet. Dieses teilte mit, dass es bereit ist, die Dateien in dieser Form als Archivalien zu bewerten und zu übernehmen, es schlug gleichzeitig vor, die Möglichkeit der Erweiterung der Anzahl der nicht anonymisierten Angaben um die Postleitzahl und den Gemeindeteil zu verhandeln. Das Statistikamt war mit dieser Erweiterung einverstanden. Der Umfang der Anonymisierung entsprach so in der Schlussphase dem ursprünglichen Entwurf des Statistikamts aus dem Jahre 2011, also noch vor dem Eingriff des Personaldatenschutzamts.

Eine weitere Frage, die es leider nicht gelungen war, zur beiderseitigen Zufriedenheit zu lösen, betraf die Form der übergebenen Dokumente. Vom Gesichtspunkt der nächsten Forschungsausnutzung bevorzugte das Nationalarchiv einen Erwerb der Daten in Datenstruktur in der Form einer Datenbank (**Abb. 1**, Seite 183) oder in der Gestalt einer Sequenzkette, statt der elektronischen Bildern der Fragebögen in einzelnen Kopien im TIFF-Format. Diese Anforderung wurde leider von Seite des Statistikamts mit der Begründung nicht akzeptiert, dass es keine Absicht des Zensusgesetzes ist, neue Dateien zum Zweck der Archivierung durch nachfolgende Bearbeitung zu schaffen.

Im Februar 2014 beendete das Statistikamt die Anonymisierung der Fragebögen, im März hat das Nationalarchiv diese als Archivalien ausgewählt und in demselben Monat wurden sie dem Nationalarchiv zur dauerhaften Erhaltung übergeben. Über 17 Millionen teilweise anonymisierte Fragebögen in der Größe von 2 TB wurden übergeben. Einzelne Fragebögen wurden auf zwei exter-

nen Festplatten übergeben und wurden in eine Adressstruktur nach dem alphabetischen Verzeichnis der Gemeinde in einzelnen Landkreisen konzentriert.

In den Jahren 2015 und 2016 hat das Personaldatenschutzamt die Kontrollen beim Statistikamt und beim Nationalarchiv durchgeführt. Die Kontrollen waren auf die Einhaltung der Verpflichtung beim Schutz der Personalangaben gezielt. Im Falle des Nationalarchivs waren sie auf die Bearbeitung der Personalangaben bei den gespeicherten Zählungsformularen aus dem Jahr 2011 gezielt. Der Abschluss war, dass weder das Statistikamt, noch das Nationalarchiv die ihnen durch das Gesetz auferlegten Verpflichtungen verletzt haben. Der Archivwert von diesen neu übernommenen Daten ist aber nicht vergleichbar mit den Zensusbeständen des 20. Jahrhundert.

Zensus 2011: Anonymisierung der Daten am Beispiel des Personenbogens - Entwürfe und Ergebnis					
	anonymisierte Daten			nicht anonymisierte Daten	
Angaben/Daten	Entwurf des Statistkams 2011	Entwurf des Personaldatenschutzamts 2011	Entwurf des Statistkams 2013	Übergebene Daten ins NA 2014	Nutzung
Wohnungsnummer					
Hausnummer					
Laufende Nr. des Gebäudes					
Kreis					
Gemeinde					
Postleitzahl					
Gemeindeteil					
Zählbezirk					
Straße					
Konskriptionsnummer					
Nachname					
Vorname					
Geburtsnummer					
Geburtsstag					
Geburtsmonat					
Geburtsjahr					
Geschlecht					
Staatsbürgerschaft					
Wohnsitz im entscheidenden Augenblick (26. 3. 2011)					
Kreis					
Gemeinde					
Konskriptionsnummer					
Gemeindeteil					
Straße					
Wohnsitz ein Jahr vor der Zählung (26. 3. 2010)					
Kreis					
Gemeinde					
Wohnsitz der Mutter zum Zeitpunkt der Geburt					
Kreis					
Gemeinde					
Familienstand					
Eingetragene Partnerschaft					
Muttersprache					
Nationalität					
Religionsgesellschaft					
Höchste abgeschlossene Ausbildung					
Fachrichtung der Ausbildung					
Anzahl der Lebendgeborenen gesamt					
Anzahl der Lebendgeborenen in der gegenwärtigen (letzten) Ehe					
Beschäftigungsverhältnis					
Beruf					
Berufliche Stellung					
Wirtschaftszweig der Beschäftigung					
Arbeits-, Schulort	tellweise anonymisiert	tellweise anonymisiert	tellweise anonymisiert	tellweise anonymisiert	tellweise anonymisiert
Verkehrsmittel					
Wegdauer					
Häufigkeit des Weges zum Arbeits- oder Schulort					

Abb. 1: Zensus 2011 – Anonymisierung

AIDS-CAMP

VALIDIERST

DU NOCH ODER

ARCHIVIERST

DU SCHON?

VALIDIERST DU NOCH ODER ARCHIVIERST DU SCHON?

Der Bedarf eines Workflow-Managements in der Formatverifikation¹

Zusammenfassung

Daniel Wittmann, Christine Träger

Im Rahmen der elektronischen Aufbewahrung und Archivierung ist eine exakte und qualitätsvolle Dateiformatidentifizierung für die gegenwärtige und zukünftige Erhaltung der elektronischen Unterlagen essentiell. In der Praxis stößt man jedoch bei der Verwendung von Software zur Erkennung und Validierung von Dateiformaten häufig an Grenzen: Keines der derzeit bestehenden Werkzeuge unterstützt alle gängigen Dateiformate, so dass allein zur Abdeckung der Formatvielfalt der Einsatz sich ergänzender Werkzeuge erforderlich ist. Die notwendige Vielfalt wird durch die Tatsache verstärkt, dass die Programme in der Regel nur mit einer der beiden etablierten Identifizierungstechniken arbeiten: Formaterkennung oder Formatvalidierung. Sollen die Identifizierungsergebnisse qualitätsgesichert werden, müssen zusätzlich zu den mindestens erforderlichen Werkzeugen weitere Erkennungs- und Validierungsprogramme hinzugezogen werden, die unabhängig von diesen Formatprüfungen durchführen und die Ergebnisse der eingesetzten Werkzeuge bestätigen bzw. zurückweisen. In der Folge müssen im Rahmen der elektronischen Aufbewahrung und Archivierung diverse Identifizierungswerkzeuge mit unterschiedlichen technischen Verfahren parallel eingesetzt werden, um einerseits eine Vielzahl verschiedener Dateiformate unterstützen zu können und andererseits eine hohe Qualität der Formatidentifikation zu erreichen.

Die verschiedenen Identifizierungsergebnisse zu einer Datei müssen während des Ingest, spätestens aber im Vorfeld einer Bestandserhaltungsmaßnahme, vom Bearbeiter gesichtet und beurteilt werden, insbesondere dann, wenn sie voneinander abweichen oder Identifizierungsprobleme aufgetreten sind. Da die Ergebnisausgaben der Werkzeuge nicht genormt sind, können diese nicht ohne weiteres maschinell abgeglichen und auch nicht ohne entsprechendes Know-how des Bearbeiters eingeschätzt werden. Die manuelle Überprüfung der Formatidentifizierungsergebnisse elektronischer Unterlagen erzeugt so reichlich Aufwand und nimmt wertvolle Zeit in Anspruch.

Digitale Archive sind heute häufig noch nicht in der Lage, mit mehreren Identifizierungswerkzeugen parallel komfortabel umzugehen, die Ergebnisse der einzel-

.....
1 Der Beitrag basiert auf dem gleichnamigen, zur 23. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ 2019 in Prag gehaltenen Camp-Vortrag (online verfügbar unter: <https://www.sg.ch/kultur/staatsarchiv/auds/2019.html> [Abruf 26.06.2019]).

nen Tools automatisiert zu normalisieren und abzugleichen und so den Auswertungsaufwand für den Bearbeiter zu minimieren. Auch aus diesem Grund werden in der Praxis häufig noch nicht ausreichend Werkzeuge nebeneinander eingesetzt. Die Prozesssteuerung anhand der Erkennungsergebnisse, also bspw. die automatisierte, gezielte Zuordnung einer Datei zu einem Validierungsprogramm auf Basis der Ersterkennung des Dateiformats, ist ebenfalls ein Desiderat.²

Im Folgenden werden, nach einer Übersicht zur derzeitigen Marktsituation der Formatidentifizierungswerkzeuge sowie der Darstellung der Umsetzung im Thüringischen Elektronischen Magazin (ThELMA)³, anhand von Beispielen die Ergebnisausgaben von Erkennungs- und Validierungswerkzeugen hinsichtlich ihrer Automationsmöglichkeiten analysiert und ein mögliches Workflow-Management für die Formatverifikation zur Diskussion gestellt.

Aktuelle Marktsituation

Es gibt Formatidentifizierungstools kommerzieller Anbieter, aber auch zahlreiche Open Source-Werkzeuge, die von der Community (bspw. über die Open Preservation Foundation⁴) weiterentwickelt werden. Die Programme lassen sich entsprechend ihrem technischen Verfahren als Werkzeuge zur Formaterkennung oder zur Formatvalidierung klassifizieren. Erkennungsprogramme, wie bspw. Apache Tika⁵ (nf. Tika) oder DROID⁶, unterstützen aufgrund ihrer formatunspezifischen Vorgehensweise ein großes Portfolio von Dateiformaten. Sie ermitteln das Format einer Datei mit Hilfe der Dateinamenserweiterung (file extension), einer so genannten magischen Zahl im Quellcode oder anhand anderer in die Datei integrierter Metadaten. Insbesondere die Verwendung der Dateinamenserweiterung ist fehlerträchtig, da diese einfach und beliebig geändert werden kann und lediglich die Formatfamilie benennt (bspw. PDF), nicht aber das konkrete Format bzw. die Version. Doch auch die anderen aufgeführten Erkennungsmethoden können hinsichtlich des Dateiformats höchstens Anhaltspunkte geben, da sich bspw. auch Metadaten, die in einer Datei integriert vorliegen, beliebig löschen oder verändern lassen.

Formatvalidierungsprogramme dagegen überprüfen eine Datei auf Konfor-

-
- 2 Claire Röthlisberger-Jourdan hat bereits 2012 in einem sehr informativen und gut verständlichen Tagungsbeitrag darauf hingewiesen, dass der Bedarf nach maschineller Unterstützung bei der Formatidentifikation spätestens mit der Zunahme der Datenmenge ansteigen wird und insbesondere die automatische Ansteuerung von Validierungswerkzeugen anhand des Ergebnisses der Formaterkennung sinnvoll ist. Vgl. dazu Röthlisberger-Jourdan, Claire: Formaterkennung und Formatvalidierung. Theorie und Praxis. In: Keitel, Christian; Naumann, Kai (Hg.): Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ und nestor-Workshop „Koordinerungsstellen“. Stuttgart 2013. S. 193–209, hier: S. 199.
 - 3 Vgl. zum Stand des Projektes Digitales Magazin des Freistaats Thüringen den Beitrag von Svenia Pohlkamp und Konrad Meckel im vorliegenden Band.
 - 4 Vgl. für weiterführende Informationen <https://openpreservation.org/> [Abruf 4. 10. 2019].
 - 5 Vgl. <https://tika.apache.org/> [Abruf 4. 10. 2019].
 - 6 Vgl. <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> [Abruf 4. 10. 2019].

mität mit allen Bestimmungen einer konkreten Formatspezifikation. So können selbst feinste Abweichungen vom Soll im Quellcode einer Datei festgestellt werden. Validationswerkzeuge sind infolgedessen häufig auf ein Dateiformat bzw. eine Formatfamilie (bspw. PDF) spezialisiert, da sie Details eines konkreten Dateiformats, also formatspezifisch, prüfen. Entsprechende Werkzeuge stehen aktuell vor allem für PDF und PDF/A, einige Auszeichnungssprachen sowie Bild- und Audioformate zur Verfügung.⁷

Formaterkennungs- und -validierungswerkzeuge ergänzen sich folglich. Selbst wenn für alle denkbaren Dateiformate Validierungsprogramme zur Verfügung stehen würden, könnte die Erstidentifizierung der ingestierten Dateien mit Hilfe eines Erkennungsprogramms und die darauf basierende Auswahl des entsprechenden Validierungswerkzeuges die Prozessauslastung und den Auswertungsaufwand reduzieren, da nicht alle Werkzeuge alle Dateien prüfen müssten.

Vereinzelt gibt es bereits Bestrebungen, mehrere Identifizierungswerkzeuge in einem Tool-Set zu vereinigen und gebündelt anzusprechen. Erfolgreiche Beispiele solcher Tool-Sets sind die Programme FITS⁸ (File Information Tool-Set) und KOST-Val⁹. FITS, eine Entwicklung der Harvard Library, setzt in der aktuellen Version 12 unveränderte Tools anderer Hersteller zur Erkennung und Validierung von Dateiformaten sowie zur Extraktion von technischen Metadaten ein; drei weitere Tools werden derzeit als Ergänzung des Tool-Sets diskutiert.¹⁰ Im Tool-Set wird jedes integrierte Programm autark und unabhängig von den übrigen eingesetzten Werkzeugen ausgeführt. Die Ergebnisdateien der einzelnen Tools werden jedoch in FITS-XML-Dateien überführt, die für den Anwender in eine gemeinsame Ergebnisdatei zusammengefasst werden. Diese Ergebnisdatei enthält somit die gefilterten Ausgaben der eingesetzten Werkzeuge in einer Übersicht – ein erster Schritt hin zu mehr Komfort und Zeitersparnis für den Bearbeiter. Das Programm KOST-Val wurde von der Schweizer Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) zur Validierung des Aufbaus und Inhalts von TIFF-, SIARD-, PDF/A-, JP2- und JPEG-Dateien sowie Submission Information Packages (SIP) entwickelt. Die Software setzt mit JHOVE¹¹, Callas pdfaPilot¹² (nf. Callas), 3-Heights™ PDF Validator von PDF-Tools¹³, jpylyzer¹⁴ und Bad Peggy¹⁵ sowohl kommerzielle als auch Open Source-Werkzeuge zur Formatidentifizierung ein. KOST-Val bietet sogar für die PDF/A-Validierung eine Qualitätssicherung durch ein zweites Werkzeug an (duale Validierung). Die Ergebnisausgabe erfolgt in

7 Als Nachschlagewerk u. a. für Formaterkennung- und -validierungstools hat sich die Community Owned digital Preservation Tool Registry (COPTR) etabliert. Für nähere Informationen vgl. http://coptr.digipres.org/Main_Page [Abruf 4. 10. 2019].

8 Vgl. <https://projects.iq.harvard.edu/fits/home> [Abruf 4. 10. 2019].

9 Vgl. https://kost-ceco.ch/cms/kost_val_de.html [Abruf 4. 10. 2019].

10 Vgl. <https://projects.iq.harvard.edu/fits/tools> [Abruf 4. 10. 2019].

11 Vgl. <http://jhove.openpreservation.org/> [Abruf 4. 10. 2019].

12 Vgl. <https://www.callassoftware.com/de/produkte/pdfapilot> [Abruf 4. 10. 2019].

13 Vgl. <https://www.pdf-tools.com/pdf20/de/produkte/pdf-converter-validation/pdf-validator/> [Abruf 4. 10. 2019].

14 Vgl. <http://jpylyzer.openpreservation.org/> [Abruf 4. 10. 2019].

15 Vgl. <https://www.coderslagoon.com/> [Abruf 4. 10. 2019].

Form einer optisch aufbereiteten Log-Datei werkzeug-übergreifend. Sie gibt an, welche Dateien eines Pakets nicht validiert wurden (weil kein Prüfmodul für das Format zur Verfügung steht) und listet die Spezifikationsverletzungen invalider Dateien detailliert auf.

Leider eignen sich Tool-Sets häufig nicht zum Einsatz in geschlossenen Softwarelösungen wie Digitalen Archiven, da die Anbindung von Einzelwerkzeugen flexibler ist und die originären Ergebnisausgaben der Werkzeuge geeigneter auf die Bedürfnisse der gewünschten Weiterverarbeitung abgestimmt werden können.

Im Thüringischen Elektronischen Magazin wurde in Zusammenarbeit mit dem Softwarehersteller H&T Greenline eine ähnliche erste Vereinfachung des Identifizierungsworkflows versucht: Während des Ingest wird zunächst eine grobe Formaterkennung mit den Werkzeugen DROID und Tika durchgeführt. Aus dem MIME-Type von Tika und der PUID von DROID wird für jede Datei ein anwendungsinternes Format gebildet¹⁶, das anschließend der automatisierten Auswahl des Validierungswerkzeuges (derzeit JHOVE und Callas) zugrunde gelegt wird. Die einzelnen Ergebnisse der eingesetzten Werkzeuge werden dem Bearbeiter im Ingestprozess gebündelt auf einer Übersichtsseite angezeigt (s. Abb. 1, Seite 195).¹⁷ In den Ereignissen wird das Ergebnis eines Tools zu einer spezifischen Datei in einem eigenen Eintrag festgehalten. Bei modularisierten Tools wie JHOVE wird auch in den Ereignissen der ThELMA-Übersichtsseite dokumentiert, wenn für eine Datei kein geeignetes Prüfmodul vorhanden ist. Zusätzlich werden die Einzelergebnisse im so genannten Validationsergebnis kumuliert. Um ein Objekt als gültig deklarieren zu können, müssen alle Prüfungen erfolgreich abgeschlossen worden sein, darunter neben den Formatprüfungen auch die Prüfung der Dateistruktur und der Nachricht selbst. Werden Fehler in der Formatidentifizierung festgestellt, sind diese in den Objekteignissen ebenfalls einzeln aufgeführt (s. Abb. 2, Seite 196). ThELMA gibt dabei auch an, wenn die MIME-Types der beiden derzeit eingesetzten Formaterkennungswerkzeuge zu einem Format voneinander abweichen und kein weiteres (Validierungs-)Werkzeug zur Verifikation zur Verfügung steht. Der Abgleich der MIME-Types wird mittels eines einfachen String-Vergleichs realisiert. Weitere Teile der Ergebnisausgaben der eingesetzten Werkzeuge werden derzeit in ThELMA noch nicht automatisiert verglichen und für den weiteren Prozess genutzt.

Wie im Folgenden gezeigt wird, ist das Mapping der vollständigen Ergebnisstrings der Werkzeuge (nicht nur des MIME-Types, sondern ggf. weiterer Elemente, wie bspw. Formatname und -version) in einen einheitlichen Ergebnisstring hinsichtlich der Abbildung des erkannten bzw. validierten Dateiformats genauer und so für die Verwendung in der weiteren Prozesssteuerung geeigneter. Nur so ist es auch möglich, das finale, eindeutige Dateiformat auszugeben, welches von den einzelnen Werkzeugen schließlich ermittelt wurde. Darüber hinaus müssen für eine Vereinfachung des Identifizierungsprozesses für den Bearbeiter Regeln

.....
16 Der Entscheidung des Softwareherstellers H&T Greenline, das interne Format in ThELMA aus diesen beiden Angaben zu bilden, basiert darauf, dass Tika den Datentyp (MIME-Type) granularer ausgibt als DROID und DROID wiederum mit dem PRONOM Persistent Unique Identifier (PUID) einen konkreten Dateityp inkl. der Formatversion zuordnet, den Tika nicht ermittelt.

17 Die originären XML-Ergebnisdateien der eingesetzten Erkennungs- und Validierungswerkzeuge werden im Archivierungspaket mitgeführt und können so bei Bedarf auch zu einem späteren Zeitpunkt aufgerufen werden.

zum Umgang mit unterschiedlichen Erkennungs- und Validierungsergebnissen der Werkzeuge festgelegt und implementiert werden.

Ausgabebeispiele von Erkennungs- und Validierungswerkzeugen

Um Anforderungen an ein Workflowmanagement für die Formatidentifizierung formulieren zu können, muss zunächst der Inhalt der Ergebnisdateien der eingesetzten Werkzeuge genauer analysiert werden. Als Beispiele für Formaterkennungsprogramme werden im Folgenden die Werkzeuge DROID und Tika herangezogen, für Formatvalidierungswerkzeuge JHOVE und Callas. Zur Analyse wurde für jedes Werkzeug die gleiche valide PDF/A-1b-Datei verwendet.

Im ersten Schritt werden die Ergebnisdateien der Erkennungswerkzeuge analysiert. Die Ergebnisdateien werden im XML-Format ausgegeben. XML-Dateien enthalten Elemente, welche durch Tags „< >“ gekennzeichnet sind. Jedes Element selbst enthält ein Attribut mit Attributname und Attributwert. Für die angegebene Beispieldatei gibt DROID ein Element mit dem Attribut „mime-type“ aus, das den Wert „application/pdf“ besitzt (s. Abb. 3, Seite 197). Dass es sich um ein PDF/A und um welche Version des Formats es sich handelt, ist jedoch nur zusammen mit den Attributen „name“ und „version“ erkennbar. Die von DROID ermittelte PUID zum Dateiformat wird in die Betrachtung nicht einbezogen, da sie aufgrund der spezifischen Ausrichtung auf die PRONOM-Datenbank, die nur DROID verwendet, für den Vergleich der Ergebnisausgaben mit anderen Tools nicht herangezogen werden kann.

In der Ergebnisdatei von Tika wird das Format der Beispieldatei, wie aus Abb. 3 hervorgeht, von mehreren Attributen angegeben. Für die Information des MIME-Type wird das DublinCore-Attribut „dc:format“ verwendet. Das Attribut ist zweimal mit jeweils zwei Werten aufgeführt: (1) „application/pdf“ „1.4“ und (2) „application/pdf“ „A-1b“. Beide Angaben sind gerechtfertigt, da ein PDF/A-1b auf ein PDF der Version 1.4 aufbaut.

Beide Werkzeuge kommen demnach zum gleichen Ergebnis, die Ergebnisdateien sind jedoch unterschiedlich aufgebaut und der Informationsgehalt der einzelnen Elemente ist nicht gleich aussagekräftig. Um die beiden Ergebnisse automatisiert miteinander vergleichen bzw. abgleichen zu können, muss das Ergebnis syntaktisch gleich aufgebaut sein. Um dies zu erreichen, ist ein Mapping der jeweiligen Ergebnisausgaben notwendig. Dazu müssen die zu verwendenden Attribute mit dazugehörigem Wertebereich (mit vereinheitlichtem Vokabular) definiert werden, um das eigentliche Mapping vornehmen zu können. Das Mapping erzeugt so für jedes Werkzeug eine neue Ergebnisdatei mit den Attributen „mime-type“ und „format-version“ (s. Abb. 3, Seite 197). Dadurch können die Ergebnisse der Werkzeuge Zeichen für Zeichen miteinander abgeglichen werden und im besten Fall – wenn zwischen den Werkzeugen Einigkeit über das erkannte Dateiformat besteht – ein Gesamtergebnis der Formaterkennung gebildet werden.

Im zweiten Schritt werden die Ergebnisdateien der Validierungswerkzeuge analysiert. In der Ergebnisdatei von JHOVE ist angegeben, welches Modul für die Prüfung verwendet wurde (s. Abb. 4, Seite 197). Darüber hinaus finden sich Informationen zum validierten Dateiformat. Das Validationsergebnis wird durch

das Attribut „status“ beschrieben. Das Vokabular des dazugehörigen Wertes umfasst die Ergebnisse „Well-Formed“, „Well-Formed and valid“ und „invalid“.

Das Validierungswerkzeug Callas hingegen drückt das Validationsergebnis in seiner Ergebnisdatei durch die Attribute „display_name“ und „display_comment“ aus. Wie aus Abb. 4 hervorgeht, wird das Validationsergebnis in Satzform beschrieben. Um die Validationsergebnisse von JHOVE und Callas nun automatisiert vergleichen zu können, muss für die verschiedenen Ausprägungen, die der Validationsstatus annehmen kann, ein einheitliches Vokabular gefunden werden. Den wenigsten Aufwand würde eine Umsetzung mit dem Vokabular „valid“ und „invalid“ erzeugen. Die Ergebnisstrings der einzelnen Werkzeuge müssen interpretiert und dem neuen Vokabular zugeordnet werden. Anschließend kann ein Mapping analog zu den Formaterkennungswerkzeugen erfolgen.

Konstruktion eines Workflows zur automatisierten Formatverifikation

Für die Konstruktion des Workflows werden die Werkzeuge nach ihrem eingesetzten Verfahren (Formaterkennung und Formatvalidierung) aufgeteilt. Es wird eine Hintereinanderausführung der Werkzeuge angestrebt. Die vorhandenen Erkennungswerkzeuge identifizieren jedes für sich das Format der jeweiligen Datei; die Ergebnisse werden im Anschluss automatisiert verglichen. Anhand des konsolidierten Ergebnisses zu einer Datei wird entsprechend dem erkannten Dateiformat gezielt ein geeignetes Validierungswerkzeug angesprochen. Sofern möglich, wird mindestens ein zweiter Validator zur Qualitätssicherung eingesetzt. Die Ergebnisse beider Validatoren werden ebenfalls verglichen, um ein Gesamtergebnis zu bilden (s. Abb. 5, erstes Beispiel, Seite 198).

Nachfolgend werden weitere Anwendungsfälle betrachtet, die verdeutlichen, wie sich der Workflow verhält, wenn die einzelnen Werkzeuge zu unterschiedlichen Ergebnissen kommen oder kein Ergebnis liefern können. Im zweiten Fall der Abb. 5 kommen die eingesetzten Werkzeuge zur Formaterkennung zu einem unterschiedlichen Ergebnis. Anhand der beiden erkannten Formate für die Datei werden zwei Validierungswerkzeuge ausgewählt und angesprochen, d.h. die Datei wird gegen jedes scheinbar erkannte Format validiert. Das Format, das ein gültiges Validationsergebnis erzielt, wird als Validationsergebnis vermerkt, da im Allgemeinen der Formatvalidation eine höhere Bedeutung zugemessen wird als der Formaterkennung. Im dritten Fall ermitteln die Erkennungswerkzeuge für eine Datei das gleiche Format, das basierend darauf gewählte Validierungswerkzeug sowie das Werkzeug zur Qualitätssicherung der Validation kommen jedoch zu jeweils unterschiedlichen Ergebnissen. In der Folge kann das Ergebnis nicht weiter automatisiert verarbeitet werden; ein manueller Eingriff der fachlichen Administration bzw. Steuerungsstelle ist erforderlich, um das Ergebnis zu interpretieren und den Ingestprozess weiterführen zu können. Ggf. bleibt das tatsächliche Dateiformat ungewiss bis eine Reidentifizierung¹⁸ durch ein weiterentwickeltes Validationswerkzeug erfolgen kann.

.....
 18 Reidentifizierung meint im Kontext der elektronischen Archivierung das beliebig häufige, erneute bzw. präzisere Identifizieren eines Dateiformats nach dessen Ingestierung und Speicherung, aufgrund der stetigen (technischen) Weiterentwicklung der Erkennungs- und Validierungsmöglichkeiten.

Im vierten Anwendungsfall liefern sowohl die Erkennungs- als auch die Validierungsprogramme kein eindeutiges Ergebnis. Es wird zwar analog zum zweiten Fall ebenfalls für jedes scheinbar erkannte Format eine Validation angestoßen. Wenn jedoch auch diese kein eindeutiges Ergebnis liefert, da z.B. die Datei für mehrere erkannte Formate gültig geprüft wurde, muss auch hier die Steuerungsstelle manuell eingreifen.

Im letzten zu betrachtenden Fall kann bereits im Rahmen der Formaterkennung kein Dateiformat ermittelt werden, gegen das validiert werden könnte. So muss der Workflow unterbrochen werden bzw. ist nicht durchführungsfähig. Die Entscheidung über das weitere Vorgehen muss von der Steuerungsstelle getroffen werden.

Ein wichtiges Kernstück des Workflows ist die Zuordnung der Formaterkennungsergebnisse zu Validierungswerkzeugen. Ein Dateiformat kann dabei mehreren Werkzeugen zugeordnet werden, wenn das eingesetzte Werkzeug in der Lage ist, mehrere Formate zu validieren. Für die Initialisierung des Workflows muss eine Tabelle erstellt werden, die die Zuordnung vornimmt. Dies ist mit einem erheblichen Erstaufwand verbunden, der zusätzlich einen regelmäßigen Pflegeaufwand nach sich zieht, insbesondere wenn neue Werkzeuge in den Workflow eingebunden werden, Updates erforderlich sind oder neue Formate hinzukommen. Auch wird in der Tabelle vermerkt, wenn zu einem Dateiformat noch kein Validator verfügbar ist. Alternativ sollte ein Default-Wert eingestellt werden können, wenn ein Format nicht in der Tabelle zugeordnet wurde, um das Format für eine Reidentifizierung vormerken zu können.

Resümee des Camps

In den letzten Jahren wurden im Arbeitskreis „Archivierung von Unterlagen aus digitalen Systemen“ immer häufiger technische Detailfragen der elektronischen Archivierung vorgestellt und diskutiert. 2017 gab es erstmals auf der 21. Tagung in Basel einen eigenen Block zum Thema Dateiformate. Das Camp F zum Bedarf eines Workflow-Managements in der Formatverifikation folgte auf der diesjährigen Tagung dieser Entwicklung und richtete sich gleichzeitig an alle interessierten Teilnehmer. Es bestand vor allem der Wunsch, Rückmeldungen dazu einzuholen und Gespräche darüber anzustoßen, inwieweit der Bedarf an einer stärkeren Automation in der Formatidentifizierung in den Archiven, Bibliotheken und anderen Einrichtungen überhaupt vorhanden ist, insbesondere vor dem Hintergrund der heterogenen Systemlandschaft des deutschsprachigen Raums, bestehend sowohl aus Eigenentwicklungen als auch aus kommerziellen Lösungen zur elektronischen Archivierung.

Im Anschluss an den Vortrag fand zunächst mit den anwesenden Vertretern der Dienstleister Startext und LVR-InfoKom ein Austausch über die Umsetzung in ihren Softwarelösungen statt. Bislang ist keine OAIS-konforme Softwarelösung bekannt, die die Formatidentifizierung in der beschriebenen Weise bereits anbietet. Die Tatsache, dass ein solches Workflow-Management von jedem Dienstleister bzw. Archivverbund aufgrund der jeweils unterschiedlichen Systembedingungen selbst entwickelt werden müsste, wurde von den Teilnehmenden des Camps als schwierig bewertet. Schließlich kam die Frage auf, wie häufig es in der Praxis vorkommt, dass ein Dateiformat etwas anderes zu sein vorgibt,

als es ist bzw. dass ein Dateiformat von den eingesetzten Identifizierungswerkzeugen falsch erkannt wird oder sich Werkzeuge im Erkennungsergebnis uneinig sind, denn nur unter diesen Umständen wäre der Aufwand für ein solches Workflow-Management gerechtfertigt. Sowohl bei den Vortragenden als auch bei den Teilnehmenden gibt es dazu noch keine verlässlichen Erfahrungswerte. Das Projektteam des Landesarchivs Thüringen berichtete, dass zumindest unterschiedliche Erkennungs- und Validierungsergebnisse der eingesetzten Werkzeuge gängig sind. Es wurde festgehalten, dass es neben den technischen Schwierigkeiten, die die Dateien selbst mitbringen können, auch Werkzeuge geben kann, die ein Dateiformat nicht zuverlässig erkennen – auch in diesem Fall ist bspw. eine Qualitätssicherung durch ein zusätzliches Werkzeug sinnvoll. Eine Teilnehmerin des Camps warf letztlich ein, dass mögliche Identifizierungsprobleme im Ingest akzeptiert werden müssten, da die übernommenen Unterlagen wohl kaum abgelehnt würden, wenn technische Probleme aufträten. Dies bestätigten die Vortragenden, allerdings mit der Einschränkung, dass in diesem Fall bei der abgebenden Stelle nachgefragt werden könnte, ob die betroffene Datei bspw. noch in einer anderen Version vorliegt oder Kenntnisse zur Datei beim Sachbearbeiter bestehen etc.

Im Ergebnis der Diskussion wurde der starke Know-how-Aufbau des Projektteams am Landesarchiv Thüringen im Bereich der Formaterkennung und -validierung von den Teilnehmenden sehr positiv bewertet. Viele Einrichtungen beschäftigen sich jedoch aktuell schwerpunktmäßig noch mit anderen Themenbereichen der elektronischen Archivierung, haben bislang teilweise keine tieferen Kenntnisse der Identifizierungswerkzeuge und dokumentieren die Formatangaben, die sie im Ingest erfassen können, für spätere Bestandserhaltungsmaßnahmen, ohne diese zwangsläufig bereits zu bewerten. Vermutlich wird erst die Archivierungspraxis der kommenden Jahre und der stetige Zuwachs an elektronischen Unterlagen zeigen, ob ein Workflow-Management im der Formatverifikation benötigt wird.

Summary

None of the existing tools for checking and verifying of file formats supports all of the common file formats. Additionally, the tools frequently use only one of the possible techniques for identification of file formats (detection or validation). In the context of digital archiving, this requires the parallel application of a wide range of different tools ensuring both the quality of file format identification and covering the whole range of file formats. Usually, the outputs' evaluation as well as the resulting continuation of the process have to be done manually. This paper shows the challenges and limits of the current practice of file format identification. It points out prospects in using workflow management to maintain the identification's quality and to (partially) automate the entire process. Initially, the tools' output files were analysed and statements identified, which are suitable for mapping the separate results into a single comprehensive identification result. Subsequently, for the purposes of building the workflow, the use cases of file format identification are discussed, starting with the event of a conclusive validation result for a file. This is followed by the event of an inconsistent output from the tools for detection and validation and, finally, the case of no fit file format identification at all.

Bildunterschriften:

Validationsergebnis für Nachricht

Übernahmepakete

Validationsergebnis ? **gültig**

Validationsergebnisse ungültiger Objekte ? **Kein(e) Übernahmeelement vorhanden.**

Unzulässige Archivierungsformate ? **Kein(e) Übernahmeelement vorhanden.**

Ereignisse ?

Erstellungszeitpunkt	Schweregrad	Operation	Nachrichtentext
22.08.2019 18:23:44	Information	Nachricht Überprüfung	Die Nachricht ist gültig.
22.08.2019 18:23:44	Information	Nachricht Überprüfung	Die Mime-Typen der Nachrichtendaten sind gültig.
22.08.2019 18:23:44	Information	Nachricht Überprüfung	Die Nachricht ist gültig.
22.08.2019 18:23:44	Information	Nachricht Überprüfung	Die Mime-Typen der Nachrichtendaten sind gültig.
22.08.2019 18:23:44	Information	Nachricht Überprüfung	Die Dateistruktur der Nachricht ist gültig.
22.08.2019 18:09:36	Information	Nachricht Erstellung	Nachricht aus Upload mit GUID: cc57a8ff-94af-4c57-9aa4-4600c36c73a9 erstellt.

6 Elemente : Anzeigen Elemente

Objekt Container

Containerereignisse ?

Erstellungszeitpunkt	Schweregrad	Operation	Nachrichtentext
22.08.2019 18:23:35	Warnung	JHOVE Metadaten Extraktion	kein JHOVE-Modul für die Datei: keine_Sonderzeichen.txt gefunden.
22.08.2019 18:23:44	Information	CeIIas PDF/A Überprüfung	PDF/A Datei: test_pdf_a2-a.pdf ist gültig.
22.08.2019 18:23:42	Information	CeIIas PDF/A Überprüfung	PDF/A Datei: test_pdf_a2-b.pdf ist gültig.
22.08.2019 18:23:40	Information	CeIIas PDF/A Überprüfung	PDF/A Datei: test_a1-a-word.pdf ist gültig.
22.08.2019 18:23:39	Information	CeIIas PDF/A Überprüfung	PDF/A Datei: test_pdf_a2-u.pdf ist gültig.
22.08.2019 18:23:37	Information	CeIIas PDF/A Überprüfung	PDF/A Datei: test_pdf_a1-b.pdf ist gültig.
22.08.2019 18:23:35	Information	Metadaten Extraktion	Metadaten aller Dateien extrahiert.

7 Elemente : Anzeigen Elemente

Abb. 1: Auszug aus der ThELMA-Übersichtsseite zu einem gültigen SIP während des Ingest.

Validationsergebnis für Nachricht

Übernahmepakete

Validationsergebnis **ungültig**

Validationsergebnisse ungültiger Objekte Zurück **1** 2 3 4 Weiter

Bezeichnung	Pfad
A_G0315_invalide.tif	/A_G0315_invalide.tif
A_G0314_invalide.tif	/A_G0314_invalide.tif
A_G0313_invalide.tif	/A_G0313_invalide.tif
A_G0312_invalide.tif	/A_G0312_invalide.tif
A_G0311_invalide.tif	/A_G0311_invalide.tif

17 Elemente : Anzeigen | 5 Elemente Zurück **1** 2 3 4 Weiter

Unzulässige Archivierungsformate Kein(e) Übernahmeelement vorhanden.

Ereignisse Zurück 1 **2** Weiter

Erstellungszeitpunkt	Schweregrad	Operation	Nachrichtentext
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	Die MIME-Typen sind ungleich und es sind keine weiteren Validationswerkzeuge verfügbar für die Datei: A_G0310_invalide.tif
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	JHOVE-Prüfung ist für die Datei: A_G0311_invalide.tif fehlgeschlagen.
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	JHOVE-Prüfung ist für die Datei: A_G0312_invalide.tif fehlgeschlagen.
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	JHOVE-Prüfung ist für die Datei: A_G0313_invalide.tif fehlgeschlagen.
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	JHOVE-Prüfung ist für die Datei: A_G0314_invalide.tif fehlgeschlagen.
23.08.2019 09:58:22	Fehler	Nachricht Überprüfung	JHOVE-Prüfung ist für die Datei: A_G0315_invalide.tif fehlgeschlagen.
23.08.2019 09:58:22	Information	Nachricht Überprüfung	Die Nachricht ist gültig.
23.08.2019 09:58:22	Information	Nachricht Überprüfung	Die Mime-Typen der Nachrichten Daten sind gültig.
23.08.2019 09:58:22	Information	Nachricht Überprüfung	Die Dateistruktur der Nachricht ist gültig.
22.08.2019 17:47:31	Information	Nachricht Erstellung	Nachricht aus Upload mit GUID: a177d1d5_4d6e8_43a1-bb66-5395ce3a1fc5 erstellt.

22 Elemente : Anzeigen | 12 Elemente Zurück 1 **2** Weiter

Abb. 2: Auszug aus der ThELMA-Übersichtsseite zu einem ungültigen SIP während des Ingest.

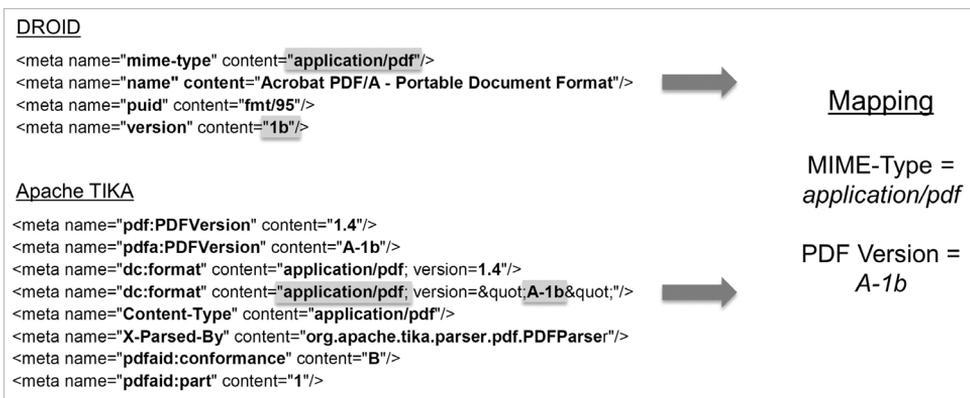


Abb. 3: Ausschnitt aus den Ergebnisdateien von DROID und Apache Tika. Die markierten Informationen werden in das Mapping übernommen.

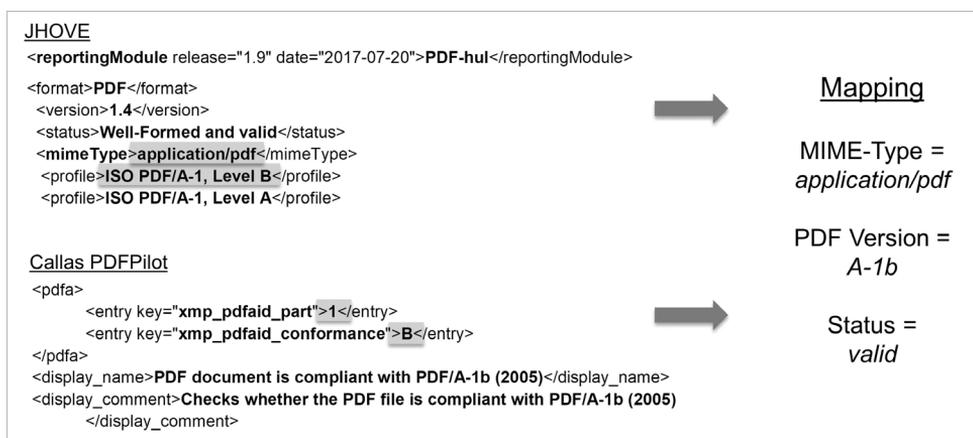


Abb. 4: Ausschnitt aus den Ergebnisdateien von JHOVE und Callas pdfaPilot. Die markierten Informationen werden in das Mapping übernommen.

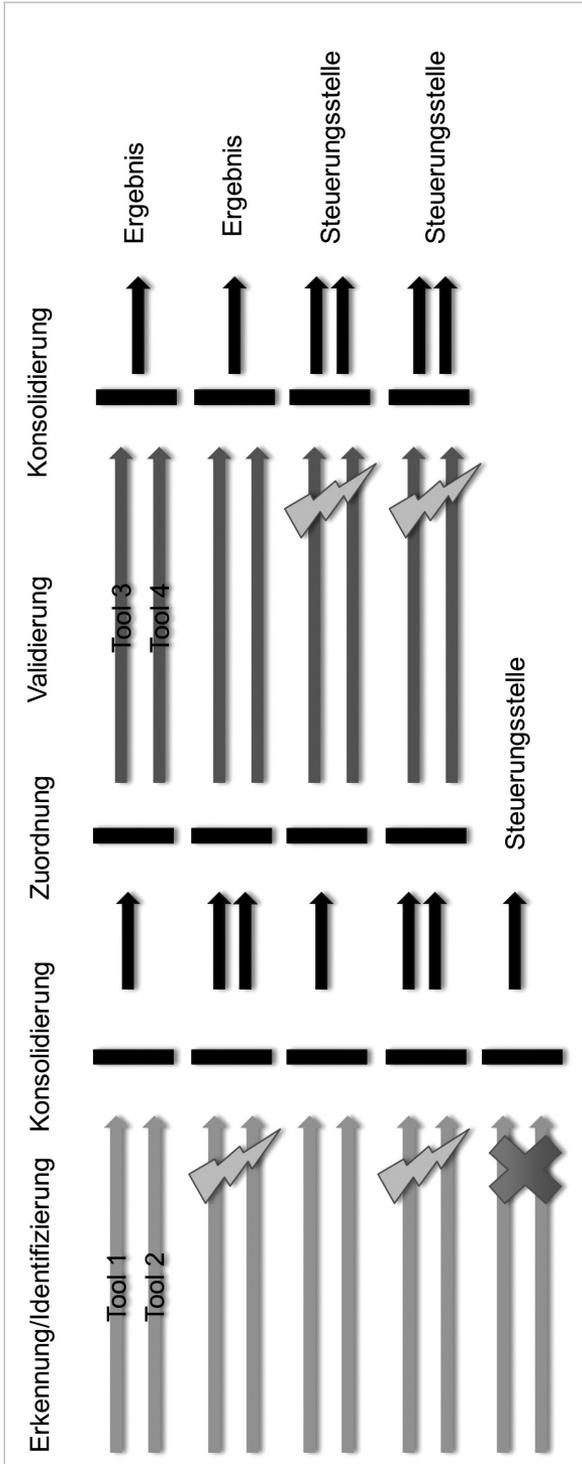


Abb. 5: Schematische Darstellung des Workflows für die Formatverifizierung und der zu berücksichtigenden Sonderfälle.

AUDS-CAMP

ARCHIVSTANDARDS

UND

BEST PRACTICES

ARCHIVSTANDARDS UND BEST PRACTICES

Christine Gigler

Zusammenfassung

Das AUdS-Camp II/G, das von drei Referent*innen bestritten wurde, stand unter dem Motto „Archivstandards und Best Practices“. Am Beginn gab es daher zunächst auch drei unterschiedliche Impulsreferate. Marta Riess von der Internationalen Atomenergiebehörde IAEA in Wien thematisierte die Herausforderungen, welche die Vielzahl an Standards, mit denen sich Archivar*innen bei der digitalen Archivierung konfrontiert sehen, mit sich bringen und welche Möglichkeiten oder Einschränkungen aus dem Spannungsfeld von Standards, Best Practices und realen Gegebenheiten erwachsen.

Anschließend widmete sich Steffen Schwalm (IT-Beratungsunternehmen „msg systems ag“) aktuellen rechtlichen Rahmenbedingungen und Standards für das Records Management und die vertrauenswürdige, beweiswerterhaltende Langzeitspeicherung geschäftsrelevanter Aufzeichnungen. Dabei gab er auch einen Überblick über den Stand der entsprechenden Standardisierungsaktivitäten auf europäischer Ebene und die aktuelle Weiterentwicklung der technischen Richtlinie BSI-TR 03125 „Beweiswerterhaltung kryptographisch signierter Dokumente“ als einem der maßgeblichen Standards auf diesem Feld.

Im letzten Impulsreferat stellte die Verfasserin dieser Zeilen, der auch die Leitung des Camps übertragen worden war, die nestor-AG „Kooperation der Archive“ und den nestor-Archivstandard vor. Der von der nestor-AG erstellte Leitfaden (nestor-materialien 22) für die Erlangung eines nestor-Archivstandards bietet ein Verfahren, anhand dessen klassische Archive ihre Anforderungen auf dem Gebiet der digitalen Archivierung formulieren und gemeinsam nach außen vertreten können.

Die sich daran knüpfende, teils lebhaft geführte Diskussion konzentrierte sich auf die Ausführungen Steffen Schwalms, dessen Thesen zur Frage, wie elektronisch signierte Dokumente über lange Zeit im Sinne eines rechtswirksamen Beweiswerterhalts vertrauenswürdiger gespeichert werden und welche Rolle den Archiven in diesem Kontext zukommen soll, von den Teilnehmenden kontrovers aufgenommen wurden.

Aufgrund der thematischen Diversität des Camps, als dessen verbindende Klammer lediglich der Begriff „Standard“ betrachtet werden kann, muss ein abschließendes Resümee entfallen.

KOMPLEXE REALITÄTEN – KOMPLEXE METADATEN

Wie man Standards, Best Practices und Realität zusammenführt

Marta Riess

Der vorliegende Beitrag war Teil eines AUdS Camps zu Archivstandards und Best Practices und wirft einige Fragen auf, die im Laufe eines Projektes zur Implementierung eines Archivmanagementsystems entstanden sind. In a nutshell kann die Hauptfrage des Vortrages zusammengefasst werden unter: wie kann man ein Metadatenmodell zwischen den besonderen Bedürfnissen des Archivs, der Compliance in der Befolgung bestimmter Standards und gleichzeitig mit ausreichender Flexibilität gestalten, um noch zukünftigen Situationen Rechnung tragen zu können?

Ausgehend von einem traditionellen Katalogisierungsprogramm entsteht die Notwendigkeit mit zunehmenden Digitalisierungsprojekten und der sich nähernden Möglichkeit, digitale Akten zu übernehmen, das – wahrscheinlich auch technisch – limitierte Katalogisierungssystem durch ein neues zu ersetzen. Der Beitrag fokussiert definitiv nicht auf den Prozess, nach welchen Kriterien man ein Archivmanagementsystem auswählen sollte oder welche Aspekte auf keinen Fall missachtet werden sollten (somit also nicht auf den Anforderungskatalog), sondern auf den Prozess, der ein Kernelement des Projektes sein wird: die Entwicklung des Metadaten-Modelles für die Applikation.¹

Metadaten, diese „Daten über Daten“, haben einen langen Weg aus der Katalogisierungs Ecke zurückgelegt und sind heutzutage ein heiß umkämpftes Feld, Informationen aus unterschiedlichen Perspektiven „greifbarer“, auch „nachvollziehbarer“ zu machen.² Die Beschreibung (das Deskriptive) steht wei-

-
- 1 Ein in diesem Sinne sehr lehrreicher Text wird von ISO selber zur Verfügung gestellt: Ohne Angaben von Autoren, beschreibt der im Internet zur Verfügung gestellte Artikel „Building a metadata schema: where to start? Unter der Nummer ISO/TC 46/SC11/N800R1 sehr genau, welche Aspekte (und ISO Standards) zur Hand genommen werden sollten: <https://committees.iso.org/files/live/sites/tc46sc11/files/documents/N800R1%20Where%20to%20start-advice%20on%20creating%20a%20metadata%20schema.pdf> [Abruf 14.07.2019].
 - 2 Ein wunderbarer Artikel, der den Stand der Diskussionen in 2014 zusammenfasste, stammt von SKINNER, Julia: Metadata in Archival and Cultural Heritage Settings: A Review of the Literature. In: Journal of Library Metadata, vol. 14, issue 1 (2014), S. 52-68; siehe auch (zwar nicht ganz so jung aber dennoch noch immer von äußerster Relevanz: WALLACE, David A.: Managing the Present: Metadata as Archival Description. In: Archivaria 39 (1995), S. 11-21; siehe auch: RILEY, Jenn: Understanding Metadata. What is Metadata and what is it for? NISO Primer (= A Primer Publication of the National Information Standards Organization), 2017 https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf [Abruf 15.07.2019]

terhin im Vordergrund, ist jedoch nur ein Aspekt unter vielen. Lassen wir all jene Diskussionen beiseite, die die Notwendigkeit von (vor allem deskriptiven) Metadaten in Zeiten von Volltextsuche und Index-crawlern überhaupt in Abrede stellen und wenden uns den unterschiedlichen Konzepten zu:

Während deskriptive Metadaten an sich wohl am wenigsten näherer Erklärung bedürfen, sind die anderen Metadaten-Typen einer näheren Betrachtung wert: strukturelle Metadaten bilden die Abhängigkeit einzelner Informationsbestandteile zueinander ab, auch sie stammen inherent aus bibliothekarischen Standards. Erweitert wurde das Konzept mit der Zunahme digitaler Inhalte, denn strukturelle Metadaten können nun auch den Typus, die Version oder die Beziehung unterschiedlicher Teile digitaler Daten zueinander beschreiben.³

Sobald man sich aus dem traditionellen bibliothekarischen Umfeld löst, werden die Schematas unklarer, die Zuordnung ändert sich je nach hinzugezogenem Schema (NISO oder MINERVA zum Beispiel): Administrative Metadaten sind laut NISO dazu da, um das Beschreibungsobjekt zu verwalten; der NISO Standard inkludiert dabei auch technische Metadaten.⁴ Das europäische MINERVA Projekt hingegen inkludiert außerdem Provenienzinformation und Rights Management.⁵ In diesem Sinne gehen MINERVA und METS Hand in Hand, wobei erstere jedoch Preservation-Metadaten von administrativen Metadaten separiert.⁶ Für technische und administrative Metadaten im Bereich der digitalen Langzeitar Archivierung steht PREMIS zur Verfügung, der Standard, der basierend auf dem OAI Referenzmodell, agent, right, event und object mit environment in Bezug zueinander stellt. Deskriptive Metadaten können quasi als ein Annex in das XML-Schema eingefügt werden (zum Beispiel als EAD).⁷

Wenn man sich dieses Potpourri an unterschiedlichen Zugängen ansieht (hier ist der Typus „administrative“ inkludiert, dort der „technische“ separiert) ist man an den ArchivarInnen scherz erinnert, dass der Versuch, die Menge der Standards zu verkleinern und sie besser zu vereinen, nur in die Erschaffung eines neuen Standards mündet. Versuchen wir uns hier auf 2 Grundprinzipien zu einigen: Metadaten dienen (auch) zum Informationsaustausch, sie sollten daher einem Schema folgen. Zweitens: in ihrem Format sollten sie inhärent durch Menschen oder Maschinen lesbar sein (das Hauptbeispiel hier wäre die Verwendung von xml).

.....
3 Siehe hinsichtlich der Klassifizierung in deskriptive, strukturelle und administrative Metadaten die Zusammenstellung durch die Cornell University: <http://preservationtutorial.library.cornell.edu/metadata/table5-1.html> [Abruf 16.07.2019].

4 NISO, Understanding Metadata, S. 6-7.

5 Technical Guidelines for Digital Cultural Content Creation Programmes (Version 1.0: Revised 08 April 2004) (<http://www.minervaeurope.org/publications/technicalguidelines/metadata.htm>) [Abruf 15.07.2019].

6 Siehe auch den Abschnitt zu administrativen Metadaten bzw. die Zusammenfassung des Standes der Diskussionen in: MATHE, Kathryn, IVACS, Gabriella: Best Practices for Trusted Digital Content Repositories, 2012, S. 48-51 (<http://www.peoplesheritage.eu/pdf/D2-4-Grant250549-HOPE-BestPracticesTrustedDigitalContentRepositories2-0.pdf>) [Abruf 15.07.2019].

7 Zu PREMIS siehe die von der Library of Congress als Hauptsponsor des Data Dictionary unterhaltene Seite: <http://www.minervaeurope.org/publications/technicalguidelines/metadata.htm> [Abruf 16.07.2019].

Steven Miller hat sehr greifbar die notwendigen fünf Schritte zusammengefasst, die notwendig sind, die Interoperabilität von Metadaten zu verbessern⁸:

- Die korrekte Verwendung eines Standards (er bezieht sich hierbei insbesondere auch Dublin Core)
- Die ausreichende Kontextualisierung der Information und die Fokussierung auf genügend Zugangspunkte zu den Metadaten (id est, ihre Interoperabilität)
- Die Verwendung von maschinenlesbaren und verlinkbaren Daten
- Die Unterscheidung zwischen administrativen, technischen und deskriptiven Metadaten
- Und schlussendlich (und nicht zu unterschätzen) Die Dokumentation des Kontextes aus welchem das Metadatenmodell entsteht

Miller schlägt bereits die Brücke zu den erweiterten Überlegungen rund um die Metadaten: nämlich den unterschiedlichen Zwecken und Aspekten, denen sie dienen sollen: Soll der Katalog ausschließlich Information zu den Beständen beinhalten oder auch das Management derselben (administrative Metadaten) dokumentieren. Somit: Was soll in diesem Katalog abgebildet werden? Was soll mit den Metadaten auch in der zumindest mittelfristigen, sonst jedoch sicher langfristigen Zukunft verwaltet werden? In welchem „größeren Ganzen“ bewegt sich der Katalog (ist das Archiv beispielsweise Teil eines größeren Verbundes und muss daher bereits einem vorgegebenen Standard folgen)? Wie sieht die „Architektur“ der unterschiedlichen Archivapplikationen aus, die vielleicht schon im Archiv in Verwendung sind? Welche Rolle wird der Katalog (und seine Metadaten) in dieser „Archiv-Architektur“ spielen?

Diese Überlegungen lassen sich somit in drei Sub-Themen teilen: 1. Analyse und Umsetzung spezieller Bedürfnisse, 2. Befolgung von Standards und 3. Flexibilität.

Zunächst wird eine Analyse der Situation des Archives, seiner Aufgaben innerhalb der Organisation/Firma/in der öffentlichen Verwaltung notwendig sein? Neben der Überlegung, ob auch das Archivmanagement abgebildet werden sollte, gilt es die „speziellen Bedürfnisse“ des Archives festzulegen: Vielleicht verwaltet das Archiv auch das Zwischenarchiv und übernimmt Aufgaben des Records Managements. Vielleicht muss das Archiv Copyrightbestimmungen besonders im Auge behalten (insbesondere zutreffend auf Photo-/Audiovisuelle Archive) oder es verwaltet verschlüsselte Informationen. Was auch immer das konkrete Szenario sein könnte, es wird immer mit Metadaten Hand in Hand gehen, die voraussichtlich nicht aus der Hand eines einzelnen Standards kommen werden... wir sind somit bereits auf dem Weg (mindestens) zur „Configuration“ eines Standards oder, anders gesagt, uns einen Standard auf unsere Bedürfnisse zuzuschneiden. Dabei wird es ein fortlaufender Prozess sein, zu entscheiden, welche Metadaten wirklich wichtig sein werden und wie weit man sich von Standards entfernen möchte.

Wie bereits vorher angesprochen wurde, wird jegliche Interoperabilität leiden, wenn existente Metadatenfelder anderswertig verwendet oder neue hinzugefügt werden. EAD (Encoded Archival Description) eröffnet bereits weitreichende Mög-

.....
 8 MILLER, Steven J.: Metadata for Digital Collections: A How-To-Do-It Manual, New York 2011, S. 245.

lichkeiten und hat mit EAD3 2015 ein neues „Update“ erfahren⁹; auch ist EAD mit ISAD-G abgeglichen, was in vielen Archiven einen entsprechend standardisierten Zugang erlaubt. Gleichzeitig wird EAD keine Records Management bezogenen Metadaten (zum Beispiel für Skartierung/Disposition oder legal Hold) sowie – et was allgemeiner formuliert – administrative Metadaten inkludieren.

Das unten exemplarisch angeführte Metadatenmodell basiert auf einer Kombination von Records in Contexts¹⁰, ISO 15489-1:2016¹¹, ISO 23081-1:2017¹², EAD und natürlich ISAD-G; in geringerem Ausmaß spielten ISO 16175-1, 2 und 3 und 30301 (aus 2011, inzwischen ersetzt wurde eine neuere Version von 2019) auch eine Rolle. Während der ISO Standard 15489 hauptsächlich Metadaten im Bereich des Records Managements bereitstellt, diente ISO 23081 hauptsächlich zur Strukturierung der ereignisbezogenen Metadaten.

Die Möglichkeit, das Metadatenchema frei entwickeln zu dürfen, ist nicht allen Archiven gegeben. Operiert es bereits in einem größeren Ganzen – beispielsweise einem Verbund – oder auf der Basis einer vorgegebenen Verwaltung (siehe Gever oder Elak), wird der Metadatenkatalog bereits vorgegeben sein. Wenn dies einerseits viel analytische Arbeit erspart, so bedeutet es auch, dass die eigene Flexibilität stark eingeschränkt ist. Wenn schon Spezifika des Archivs nun nur schwer oder gar nicht beachtet werden können, so wird dies noch deutlicher werden, wenn es um die eigene „Architektur“ geht, in die das Archivmanagementsystem eingebettet werden sollte.

Das exemplarisch angeführte Schema¹³ versucht deskriptive, technische und administrative Metadaten zu verbinden, ohne die Möglichkeit auszuschließen auch digitale Akten damit verwalten zu können. Die derzeitige Ausgangssituation sind physische, hybride und digitalisierte Akten und Bestände; ob die Bestände eines digitalen Archives tatsächlich eines Tages im Archivmanagementsystem vollständig verwaltet werden, ist noch unklar aber nach derzeitigem Stand eher unwahrscheinlich. Dies hat sehr einfache Gründe: die Firma selber, von welcher das off-the-shelf Produkt gekauft wurde, sieht sich nicht als ein Katalogisierungs- und Verwaltungsprogramm für digitale Bestände im Sinne eines OAIS-Referenzmodells. Die Gefahr einer zu starken Customization überwiegt die Vision eines Archivmanagementsystems für alles.

-
- 9 Dieser von der Society of American Archivists und der Library of Congress getragene Standard kann am besten über seine Homepage angesehen werden: <https://www.loc.gov/ead/> [Abruf 4. 10. 2019].
 - 10 Records in Contexts ist ein Modell, dass seit 2016 durch der Expertengruppe zur archivarischen Beschreibung (Expert Group on Archival Description) des International Council on Archives entwickelt (und scharf kritisiert) wird. RiC can am besten über <https://www.ica.org/en/egad-ric-conceptual-model> [Abruf 4.10. 2019] erreicht werden. Eine der wichtigsten Kritiken zu RiC wurde von InterPares verfasst: <https://interparestrust.com/2016/12/11/interparestrust-responds-to-egad-ric/>.
 - 11 Es handelt sich um den überarbeiteten Standard von 2001: ISO 15489-1:2016: Information and documentation -- Records management -- Part 1: Concepts and principles.
 - 12 ISO 23081-1:2017: Information and documentation – Records management processes – Metadata for records – Part 1: Principles und ISO 23081-2:2009: Information and documentation – Managing metadata for records – Part 2: Conceptual and implementation issues.
 - 13 Das Schema wurde als Teil eines Projektes erarbeitet, dass in einem englischsprachigen Umfeld ausgeführt wurde; es ist daher auf Englisch abgefasst.

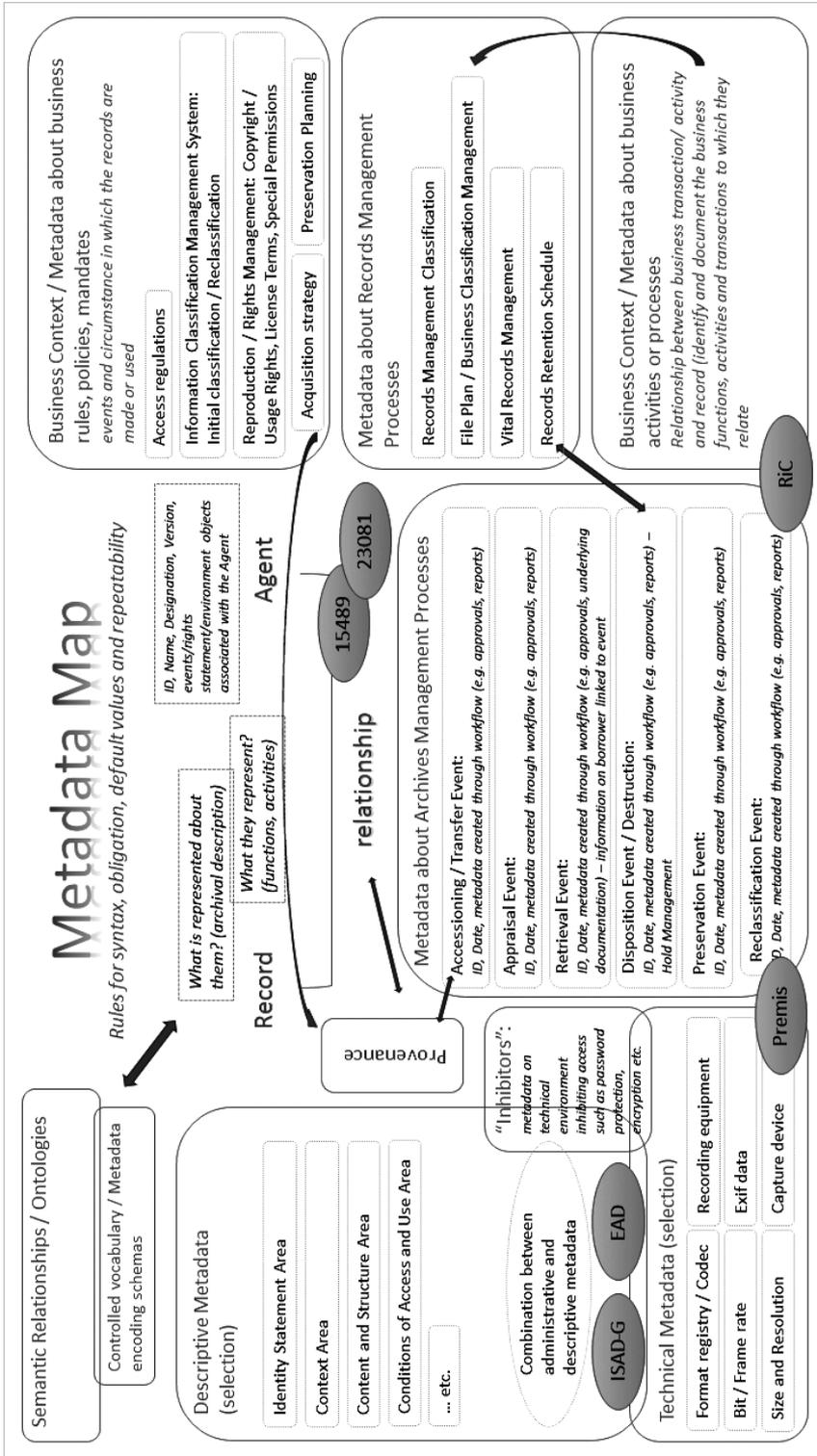


Abb. 1: Schema

Es stellt in Anlehnung an Records Management Standards das Record (die Akte), den Agent (die handelnde Identität) und die Beziehung derselben zueinander in den Mittelpunkt. Dabei wird die Akte nicht nur objektiviert (was über das Objekt ausgesagt werden kann), sondern auch subjektiviert (was das Objekt über anderes ausdrückt). Der Entscheidung, das Schema wie sichtbar zu strukturieren, gingen lange Überlegungen voraus und ist stark von der Tatsache beeinflusst, dass das Archiv auch als Zwischenarchiv fungiert und Records Management Aufgaben übernimmt. Ein Zugang rein aus der Perspektive ISAD-G erschien nicht nur nicht ausreichend, weil die hybriden und digitalisierten Bestände nicht ausreichend abgebildet werden konnten, sondern auch um auf die Ereignisse im Archivmanagement und eben die Notwendigkeiten im Zwischenarchiv hinweisen zu können. Stattdessen wurde der oben sichtbare prozessorientierte Zugang gewählt (basierend auf ISO Standard 23081):

Record und *Agent* sowie ihr *relationship* treten in Verbindung mit ihrem Business Context, das heißt, den Umständen und Ereignissen in denen sie entstanden und benutzt wurden. Hier werden auch Metadaten zu den Zugangsrechten, dem Informationssicherheits-Management und den Rechten (Copyright, Nutzungsrechte, Lizenzen oder spezielle Regelungen) Platz finden. Das Mandat oder die Grundsätze können jedoch nicht nur im operationellen Bereich der abgebenden Stellen liegen sondern auch im Archiv selber, daher spielen hier auch die Bestandsbildung und das *Preservation Planning* eine Rolle.

Records Management Metadaten sind in zwei Sub-Gruppen geteilt: jene, die RM Prozesse abbilden wie zum Beispiel der Aktenplan und die damit verbundene RM Klassifizierung, den Skartierungsplan mit Skartierungsmetadaten inklusive „*Legal Hold*“, id est dem temporär beschränkten Aussetzen des Skartierungsplanes, oder *Vital Records Management*. Und jene, die den Kontext zwischen den Records und den mit ihnen in Verbindung stehenden Geschäftsprozessen herstellen. Im Gegensatz zu der zuerst besprochenen Kategorie des Business Kontextes geht es nicht um den Makrolevel von Mandaten, sondern um den Mikrolevel von bestimmten Funktionen und Aktivitäten sowie den Kontextualisierung von transaktionalen Systemen.

Als Archivmanagementsystem fungiert die Applikation nicht nur als Katalog sondern soll auch helfen, alle Ereignisse, die mit den Objekten in Verbindung stehen, abzubilden. Darunter können Übernahme, Bewertung, Ausheben, Konservierungsmaßnahmen, Änderung der Geheimhaltungsstufe oder (wieder zurückgreifend auf die Records Management Metadaten) das Skartierungsereignis sein. Dieser ereignisbasierende Zugang basiert auf ISO 23081 erinnert jedoch durchaus auch an PREMIS. Der Aufbau des Metadatenkatalogs für Ereignisse folgt einem System: Vergabe einer ID, Datum des Ereignisses, Workflow basierende Metadaten (zum Beispiel der Genehmigung einer Skartierung oder einer Reklassifizierung).

Die Provenienz wurde als eigene Entität in dem Schema dargestellt, obwohl sie natürlich auch in ISAD-G inkludiert wäre; das Herausstreichen der Provenienz hat eher etwas damit zu tun, dass die Notwendigkeit, mehr administrative Geschichte zu betreiben, um sie korrekt und in allen Details über den Weg eines kontrollierten Vokabulars in die deskriptiven Metadaten einfügen zu können, betont werden sollte. Die Provenienz steht primär mit der Bestandsbildung (von welchen Stellen wird übernommen?, „*office of record*“), der Beziehung zwischen Record und Agent und der Übernahme ins (Zwischen)Archiv in Verbindung.

Die deskriptiven Metadaten, die in diesem Schema am meisten vernachlässigt wurden, weil sie am ehesten den Standards von ISAD-G und EAD folgen, werden noch von technischen Metadaten sekundiert, die primär der Beschreibung von Digitalisaten dienen. Für die Zusammenstellung dieses Katalogs wurden primär audiovisuelle Standards hinzugezogen, insbesondere der NISO Standard für Photos¹⁴ und die Publikationen von IASA¹⁵ und der Library of Congress¹⁶. Die Metadaten sind grob in einen Identifikationsteil (der nicht repetieren würde, was im deskriptiven Teil soundso schon ausgeführt wäre), einen Audio, einen Video, einen Audio-Video und einen Exif-Teil aufgeteilt. Die jeweiligen Abschnitte enthalten Informationen zum technischen Umfeld (*environment*), in dem sie entstanden sind, id est dem Aufnahme- oder Digitalisierungsgerät und lehnen sich hier an das PREMIS Konzept an.

Ebenfalls an PREMIS angelehnt ist das Konzept der „Inhibitors“, id est Metadaten zum technischen Umfeld (*environment*), die den Zugang unterbinden, wie Passwörter oder Kryptographie. Da die derzeit tatsächlich im Archivmanagementsystem verwalteten Digitalisate keine solche *inhibitors* aufweisen und Pläne, die Applikation auch für das Digitale Archiv zu verwenden, auf Eis gelegt sind, sind diese Metadaten nicht inkludiert worden.

Die semantischen Beziehungen oder Ontologien und Thesauern, die in dem Schema ebenfalls Erwähnung finden, dienen ausschließlich dazu, die Möglichkeit nicht außer Acht zu lassen, dass in Zukunft gerade zum Beispiel die Beschlagwortungen nicht nur auf der Basis von archivarischen kontrollierten Vokabularen sondern auch mit den in der Organisation eingeführten Ontologie für die Öffentlichkeitsarbeit verbunden werden könnte, was gerade im Sinne einer Öffnung des Katalogs für interne und externe Benutzer von Nutzen sein könnte.

Auf Basis des hier vorgestellten Schemas wurde eine konkrete Liste aller Metadatenfelder in allen ihren Details (Definition, Obligation, Wiederholbarkeit etc.) zusammengestellt und damit der tatsächliche Metadatenkatalog entwickelt. Die Analyse und die Recherche, die im Vorfeld notwendig waren, sind zeitintensiv und oftmals auch überwältigend. Unterschiedliche Standards, Konzepte und Ansätze wollen abgewogen und unter einem Hut gebracht werden, anderes kann sich im Laufe des Prozesses als Unnötig herausstellen und desto klarer sich eine Struktur herauskristallisiert, nach welcher man arbeiten möchte, desto eher gerät eines der ursprünglichen Maxime (die Beibehaltung der Flexibilität) ins Hintertreffen.

Wenn man sich die einleitenden Hauptfragen ansieht, ist dieser Text keine Anleitung, wie man ein Metadatenschema baut. Ganz im Gegenteil: es greift verschiedene Ansätze auf (Ansätze auch deshalb, weil mehr Detailismus den Rahmen der Vortragszeit gesprengt hätte) und bietet verschiedene Ideen, wie das Schema angelegt sein könnte, entweder orientiert nach Standards oder nach Prozessen (wie hier angewandt). Wie immer wieder im Vortrag hingewie-

.....
14 ANSI/NISO Z39.87-2006 (R2017) Data Dictionary - Technical Metadata for Digital Still Images (2017).

15 IASA (international Association of Sound and Audiovisual Archives) Cataloguing Rules (1999), siehe auch: <https://www.iasa-web.org/cataloguing-rules> [Abruf 15.07.2019].

16 AudioMD und VideoMD sind die XML-Schemas der Library of Congress für technische Metadaten aus digitalen Audio und Videodateien, siehe Startseite <https://www.loc.gov/standards/amdvmd/> [Abruf 16.07.2019].

sen, muss die Besonderheit eines jeden Archivs beachtet werden (es wird keinen Sinn haben, ein Schema zu erarbeiten, das schlussendlich nicht funktioniert), gleichzeitig muss man sich der Konsequenzen klar gewahr sein, wenn man sich immer weiter von Standards entfernt. Und *last but not least* sollte immer im Auge behalten werden, dass das Model flexibel genug bleibt, um für zukünftig antizipierte Situationen angepasst werden zu können.

AKTUELLE STANDARDS UND NORMEN IN RECORDS MANAGEMENT UND BEWEISSICHERER LANGZEITSPEICHERUNG

Tomasz Kusber, Steffen Schwalm, Ulrike Korte

Einführung

Die Nutzung der Informationstechnologie zur Abbildung elektronischer Geschäftsprozesse ist in Wirtschaft und Verwaltung etabliert. Der Fokus liegt dabei zunehmend auf der Umsetzung vollständig digitaler Transaktionen, Ende zu Ende, also vom Kunden zum Unternehmen/Behörde und zurück. Beschleunigt wird diese Entwicklung durch gesetzliche Vorgaben wie z. B. die Pflicht zur Einführung der elektronischen Akte und dem Onlineangebot aller digitale abbildbaren Behördenleistungen für die öffentliche Verwaltung oder die Zahlungsverkehrsrichtlinie PSD2 in der Finanzwirtschaft. In der Folge liegen geschäftsrelevante Aufzeichnungen ausschließlich elektronisch vor. Gleichzeitig sind gesetzliche Vorgaben bzgl. Zeitpunkt sowie Art und Weise der vollständigen Umsetzung digitaler Transaktionen oder elektronischer Akten einzuhalten. Eine wesentliche Kernanforderung bildet hierbei der Aufbau und Etablierung vertrauenswürdiger digitaler Transaktionen und Aufzeichnungen. Dies erfordert zum einen die eindeutige Identifizierung der beteiligten natürlichen wie juristischen Personen oder auch Verfahren (Maschine-Maschine-Kommunikation) inklusive der eindeutigen Zuweisbarkeit und Nichtabstreitbarkeit von Transaktionen und Dokumenten, die Unverändertheit geschäftsrelevanter Aufzeichnungen, zum anderen deren Verfügbarkeit in der notwendigen Form, was auch die Verkehrsfähigkeit bedingt und die Wahrung der Vertraulichkeit gegenüber Nutzern und Betroffenen. Die Maßgaben sind bis zum Ablauf der geltenden Aufbewahrungsfristen gegenüber Prüfbehörden, Gerichten, Dritten entsprechend nachzuweisen. [Ko13], [KuSc16], [Ro07], [Fi06], [We18].

Digitale Daten sind aus sich selbst heraus jedoch weder wahrnehmbar und lesbar, noch liefern sie Hinweise für ihre Integrität, Authentizität oder Ordnungsmäßigkeit im elektronischen Rechts- und Geschäftsverkehr. Vor dem Hintergrund der erwähnten Dokumentations- und Nachweispflichten, die zum einen für Fristen zwischen zwei und 110 Jahren zu erfüllen sind, die teilweise erst Jahrzehnte nach dem Abschluss des zugrundeliegenden Geschäftsvorfalles beginnen, gilt es, organisatorische und technische Maßnahmen nach dem Stand der Technik zu treffen, um die Vertrauenswürdigkeit von Prozessen und Aufzeichnungen nicht nur lang-

fristig zu gewährleisten, sondern auch nachweisen zu können. Eine langfristig vertrauenswürdige Digitalisierung kann als elementare Grundlage nachhaltiger elektronischer Prozesse, Dienstleistungen und Aufzeichnungen bezeichnet werden.

Die Verwendung kryptographischer Sicherungsmittel wie mindestens fortgeschrittener elektronischer Signaturen, Siegel und Zeitstempel ermöglicht nach geltendem Recht nicht nur die Wahrung von Authentizität und Integrität von Transaktionen und Aufzeichnungen, sondern vor allem die Erhaltung des zur Nachweisführung notwendigen Beweiswerts direkt an den aufzubewahrenden Aufzeichnungen und damit ohne deren Verkehrsfähigkeit einzuschränken (Beweiswerterhaltung). Darüber hinaus ist es notwendig, die Interpretierbarkeit der Aufzeichnungen zu gewährleisten, sie also entsprechend dem Aufbewahrungszweck und -vorgaben verfügbar zu halten (Informationserhaltung) [Ko13][Ro07], [KoKuSc18]. [EIDAS], [VDG].

Die Umsetzung erfolgt in Praxis auf Basis geltender Standards und Normen zum Records Management sowie zur beweissicheren Langzeitspeicherung, die es ermöglichen, zum einen die organisatorische wie technischen Rahmenbedingungen zum Umgang mit geschäftsrelevanten Aufzeichnungen zu etablieren und zum anderen die zur Beweiswert- und Informationserhaltung notwendigen Information, gemeinsam mit den aufzubewahrenden Aufzeichnungen in selbsttragenden Archivinformationspaketen (AIP) in einem digitalen Langzeitarchiv auf Basis [ISO15489], [ISO14721], [DIN31647], [TR03125]. Als Stand der Technik gelten Standards und Normen anerkannter wie unabhängiger Standardisierungsorganisationen wie z. B. DIN, ISO, ETSI/CEN oder BSI.

Der vorliegende Aufsatz beschreibt auf Basis der wesentlichen regulatorischen Rahmenbedingungen (Kap. 2.1) sowie fachlich-technischen Anforderungen (Kap. 2.2) einer vertrauenswürdigen Digitalisierung den aktuellen Stand der relevanten Standards und Normen (Kap. 3). Der Fokus liegt dabei, im Sinne der Nachweisfähigkeit, auf den Maßgaben zur langfristigen wie beweissicheren Aufbewahrung geschäftsrelevanter Aufzeichnungen. Darüber hinaus wird ein Ausblick auf aktuelle Anwendungsfälle sowie künftige Weiterentwicklungen im Kontext disruptiver Technologien gegeben (Kap. 4).

Regulatorischer Rahmen und Vertrauenswürdigkeit in der Digitalisierung

Regulatorischer Rahmen in Deutschland und Europa

Die eIDAS-Verordnung [EIDAS] schuf im EWR einheitliche Vorgaben für vertrauenswürdige digitale Transaktionen auf Basis elektronischer Identifizierungsmittel sowie elektronischer Vertrauensdienste im Binnenmarkt. Als Identifizierungsmittel gilt hierbei z. B. der neue Personalausweis in Deutschland sowie dessen europäische Pendant. Die Vertrauensdienste umfassen elektronische Signaturen, Siegel, Zeitstempel, Verifikationsdienste (Prüfung von Signaturen, Siegeln etc.), Bewahrungsdienste (Beweiswerterhaltung), Einschreib- und Zustelldienste (z. B. De-Mail) sowie Websitezertifikate. Durch die Zulassung z. B. von Server- und Fernsignaturen ohne Signaturkarte sowie von Siegeln (Signaturen für Organisationen) ebenso wie die Pflicht zur Anerkennung jeder mindestens fortgeschrittenen elektronischen Signatur bzw. Siegel bzw. Zeitstempel (Art. 25, 35, 41 [EIDAS]) jedes qualifizierten europäischen Vertrauensdienstes

durch öffentliche Stellen ist ein deutlicher Anstieg signierter Dokumente in Deutschland bereits im Gang, wie dies auch in den europäischen Nachbarstaaten bereits seit langen Jahren der Fall ist [VS18], [KSDV15] [KoScKu18]. Die [EIDAS] definiert für die Bewahrungsdienste gemäß Artikel 34 [EIDAS] auch spezielle Anforderungen für die beweiserhaltende Aufbewahrung

Im Zuge der Umsetzung der [EIDAS] werden „beim Erlass von delegierten Rechtsakten bzw. Durchführungsrechtsakten, die von europäischen und internationalen Normungsorganisationen und -einrichtungen, insbesondere dem Europäischen Komitee für Normung (CEN), dem Europäischen Institut für Telekommunikationsnormen (ETSI), der Internationalen Normungsorganisation (ISO) und der Internationalen Fernmeldeunion (ITU), festgelegten Normen und technischen Spezifikationen gebührend berücksichtigt“ [EIDAS]. Die genannten Standardisierungsgremien haben mit Mandat M416 der Europäischen Kommission die explizite Aufgabe, die rechtlichen Regelungen der [EIDAS] durch konkrete technische Standards untersetzen und so die Umsetzung durch Interoperabilität und Harmonisierung von Vertrauensdiensten und elektronischen Identifizierungsmitteln zu erleichtern [Ko17]. Diese technischen Standards werden zudem zunehmend weltweit angewandt, was deren Bedeutung hervorhebt [KSDV15]. Die nachstehende Grafik zeigt das Zusammenwirken von Rechtsrahmen [EIDAS] und den technischen Normen durch ETSI/CEN im Überblick.

Abb. 1, Seite 227

In Deutschland wird die wirksame Durchführung von [EIDAS] durch das Vertrauensdienstegesetz [VDG] geregelt. Dieses sieht die Beweiserhaltung signierter Dokumente in § 15 [VDG] vor. Die Umsetzung vertrauenswürdige digitaler Transaktionen sowie deren Nachweis wird durch weitere EU-Vorgaben [EU-DLR], [EUDSGVO] einerseits sowie durch branchenspezifische Vorgaben andererseits, so z. B. [EASA], [FDA], [GXP] in Luftfahrt, Gesundheitswesen und Pharma, [EuroSOX], [PSD2] in der Finanzindustrie, oder Maßgaben zum Haftungsrecht [BGB] andererseits untersetzt [Ko16], [We18].

Eine weitgehende Digitalisierung der deutschen Behörden (z. B. Einführung einer führenden E-Akte, Umsetzung vom rechtssicheren ersetzenden Scannen sowie des Onlineangebots aller Behördenleistungen oder Implementierung von beweiserhaltenden Aufbewahrungsdienste ist bis 2020 bzw. 2025 vorgesehen, der Justiz bis 2022 [EGOVG], [OZG] [eJusticeG]. Entsprechende Pendantes zum [EGOVG] liegen in nahezu allen Bundesländern vor. Die Vorgaben des [OZG] gelten für alle Verwaltungsebenen.

Untersetzt wird der regulatorische Rahmen durch verbindliche Standards und Normen, die als Stand der Technik gelten. Um der Verpflichtung nachzukommen, behördliche Entscheidungen bis zum Ablauf der geltenden Aufbewahrungsfristen nachweisbar und damit gerichtsfest zu halten sind insbesondere die Maßgaben zum Records Management sowie die Technischen Richtlinien RESISAN (rechtssicheres ersetzendes Scannen) [TR03138] und TR-ESOR [TR03125] hervorzuheben. Sowohl die TR-RESISCAN als auch die TR-ESOR sind zudem rechtlich als Stand der Technik verankert [MiniKeGovG], [We18], [Ko16], [ZPO]. Die Bedeutung eines ordnungsgemäßen Records Management sowie der Umsetzung der genannten Technischen Richtlinien des BSI unter Nutzung der Vertrauensdienste der [EIDAS] zeigt sich zudem in einigen

aktuellen Gerichtsurteilen, welche die Umsetzung des Stands der Technik faktisch einfordern [Az. 6 K 691/14.WI.A; Az. 6 K 808/17.WI.A].

Vertrauenswürdigkeit digitaler Transaktionen und Aufzeichnungen

Vertrauenswürdigkeit geschäftsrelevanter Aufzeichnungen und Transaktionen erfordert in Deutschland und Europa vertrauenswürdige Dritten, die auf Basis gesetzlicher Vorgaben sowie geltender Standard und Normen agieren und wiederum durch vertrauenswürdige Dritte in einer Vertrauenskette transparent überprüft wie ermächtigt werden. Die nachstehende Grafik zeigt diese Vertrauensketten am Beispiel der [EIDAS].

Abb. 2, Seite 228

Wesentliches Merkmal einer vertrauenswürdigen Digitalisierung ist Gewährleistung und Nachweis der zentralen Schutzziele:

- Integrität beinhaltet
- Authentizität
- Nachvollziehbarkeit
- Verfügbarkeit beinhaltet
- Lesbarkeit/Nutzbarkeit
- Verkehrsfähigkeit
- Vertraulichkeit

elektronischer Prozesse anhand der geschäftsrelevanten Aufzeichnungen auf Basis der durch vertrauenswürdige Dritte definierten regulatorischen wie fachlich-technischen Anforderungen.

Nach derzeitiger Rechtslage ist kein IT-Verfahren, Organisation oder System aus sich selbst heraus vertrauenswürdig, wie dies bspw. von öffentlichen Archiven mit dem Prinzip *ius Archivi* oder aktuell für Blockchain für sog. Public Chains postuliert wird. In jedem Fall ist der Nachweis gegenüber Gerichten, Prüfbehörden etc. zu führen. Anhand Blockchain bestehen derzeit verschiedene Forschungsprojekte sowohl zur Integration von Public Chains ins bestehende Vertrauensmodell als auch für alternative Vertrauensmodelle. Konkrete Auswirkungen auf den regulatorischen Rahmen sind aktuell nicht absehbar. [DIN-SPEC31648], [KoScKu18], [KoBeScKu18], [Ro07], [BMW07], [ZPO] [We18], [He18], [ISO15489].

Wesentlicher Stand der Technik im Überblick

Grundsatz

Wie in den vorherigen Kapiteln dargestellt, bilden Standards¹ und Normen² anerkannter Standardisierungsorganisationen den sog. Stand der Technik. Hierunter werden also die fachlich-technischen Rahmenbedingungen zur Umsetzung und Nachweis einer vertrauenswürdigen Digitalisierung entsprechend den regulatorischen Vorgaben verstanden. Welcher Standard anzuwenden ist, ergibt sich demgemäß aus der Branche und konkreten Anwendungsfall der je-

1 Standard = entwickelt von anerkanntem Standardisierungsgremium, jedoch keine formale DIN-/ETSI oder CEN/ISO-Norm

2 Norm = Standard eines formalen Normungsgremiums (DIN, ETSI/CEN, ISO)

weiligen Institution. Die nachstehende Grafik gibt einen Überblick wesentlicher Standards und Normen zum langfristigen Nachweis digitaler Transaktionen und Aufzeichnungen.

Abb. 3, Seite 229

Eine beweissichere Langzeitspeicherung gewährleistet dementsprechend sowohl die Erhaltung und Nachweis der Integrität, Authentizität und Vertraulichkeit geschäftsrelevanter Aufzeichnungen als auch deren Nachvollziehbarkeit, Verfügbarkeit und Verkehrsfähigkeit. Sie umfasst also:

- Informationserhaltung
- und
- Beweiswerterhaltung [Ko13] [KoScDH14] [KoScKu18].

Elementare Basis bildet ein ordnungsgemäßes Records Management. Das in [ISO14721] normierte OAIS-Modell sowie die hierauf aufbauenden [DIN31644] etc. definieren die notwendigen Prozesse und Informationspakete eines vertrauenswürdigen digitalen Langzeitarchivs. Die [DIN31647] wiederum beschreibt die notwendigen Funktionen zur Beweiswerterhaltung in einem OAIS-konformen vertrauenswürdigen digitalen Langzeitarchiv (dLZA). Um Authentizität und Integrität langfristig zu erhalten und nachzuweisen, gilt es, das dLZA um die notwendigen technischen Maßnahmen, also Signaturen, Zeitstempel oder Siegel auf Basis der im Zuge der [EIDAS] entstandenen ETSI- vor allem hinsichtlich der Bewahrungsdienste gem. Art. 34 [EIDAS], zu ergänzen. Hierzu gehören vor allem [ETSI TS119511] und [ETSI TS119512]. Die [TR03125] des BSI führt die Vorgaben zur Informations- und Beweiswerterhaltung logisch zusammen und definiert im Kern eine Middleware zur Beweiswerterhaltung einschließlich der notwendigen Informationspakete sowie der Maßgaben zur Integration in ein vollständiges OAIS-konformes dLZA [KoScKu18], [Ko13], [KuSc16]. Im Folgenden wird auf den Stand der Standardisierung insbesondere im Records Management sowie zur Beweiswerterhaltung näher eingegangen.

Records Management

Im Jahr 2016 wurde die ISO-15489 als zentrale Norm zum Records Management neugefasst. Der Fokus liegt nunmehr vorrangig auf elektronischen Aufzeichnungen. Die Norm reiht sich ein in den mit der [ISO30300] (Terminologie), [ISO30301] (Basisanforderungen) und [ISO30302] (Guidelines zur Implementierung) geschaffenen fachlichen Rahmen zum Records Management. Die 303xx-Familie richtet sich dabei an Management von Organisationen und beschreibt grundlegende Anforderungen an den Umgang mit geschäftsrelevanten Aufzeichnungen. Diese Vorgaben bilden die Basis für spezifische Standards zur fachlich-technischen Untersetzung. Das nachfolgende Bild zeigt das neue Standardisierungsframework zum Records Management im Überblick:

Abbildung 4: ISO-Standardisierungsframework zum Records Management, Seite 230

Die ISO-15489 selbst fordert zum einen die Definition klarer wie verbindlicher Regelungen, Prozesse, Rollen und Verantwortlichkeiten für die Erzeugung/Empfang, Ordnung, Speicher, Nutzung und Aufbewahrung – kurz das Management geschäftsrelevanter Aufzeichnungen auf Basis der jeweils geltenden regulatorischen Vorgaben. Ziel sind die Wahrung und langfristiger Nachweis der Authentizität, Zuverlässigkeit (inkl. Nachvollziehbarkeit und Vertraulichkeit),

Integrität und Nutzbarkeit (inkl. Verkehrsfähigkeit) der geschäftsrelevanten Aufzeichnungen. Speziell im Kontext der Zuverlässigkeit wird die eindeutige Zuweisbarkeit explizit hervorgehoben. Hierauf basierend beschreibt ISO-15489 Anforderungen an die Identifikation geschäftsrelevanter Aufzeichnungen, die Systematisierung wie Indexierung von Aufzeichnungen sowie deren berechtigter Nutzung im Geschäftsprozess. Hervorzuheben ist zudem der Vorgehensvorschlag zur organisationsinternen Konzeption und Implementierung eines ordnungsgemäßen Records Management einschließlich der IT-Verfahren, die in geschäftsrelevanten Aufzeichnungen entstehen, empfangen, verwaltet und aufbewahrt werden einschließlich Ansätzen für eine spätere Migration im Sinne des Langzeiterhalts der Unterlagen. Im Kern fordern ISO-30301 und 15489, dass geschäftsrelevante Aufzeichnungen als Nachweis geschäftlicher Transaktionen gegenüber Prüfbehörden, Gerichten, Dritten dienen. Um dies zu erreichen, sind neben dem regulatorischen Rahmen die Maßgaben dieser Normen sowie der hierauf aufbauenden technischen Standardisierung entsprechend zu berücksichtigen [To07] [We18], [WeSc18].

Die für die Normung zum Records Management verantwortliche Working Group 2 im DIN NID 15 hat, angesichts des klaren digitalen Fokus der Norm, bewusst gegen eine Übersetzung der ISO-15489 entschieden. Stattdessen werden die regulatorischen, fachlichen und technischen Anforderungen an ein ordnungsgemäßes Records Management sowie dessen pragmatische Umsetzung in einem fundierten Leitfaden beschrieben, der zudem die Norm selbst enthält [We18]. Untersetzt werden die ISO-/DIN-Normen durch themen- und branchenspezifische Standards wie z. B. die TR-RESISCAN des BSI (branchenübergreifend) [TRO3138] zum rechtssicheren ersetzenden Scannen, das Organisationskonzept elektronische Verwaltungsarbeit (öffentliche Verwaltung) [BMI12] oder den LOTAR im Bereich der Luft- und Raumfahrt [LOTAR].

Beweissichere Langzeitspeicherung

Bewahrungsdienste gem. eIDAS (Service-Provider)

Die Standardisierung zur beweisicherten Aufbewahrung fokussierte bislang vorrangig auf die technischen Verfahren sowie die aufzubewahrenden Aufzeichnungen selbst. Hinsichtlich der Betreiber entsprechender digitaler Langzeitarchive lag zwar mit der [ISO16363] oder der [DIN31644] Standards mit entsprechenden Zertifizierungsverfahren vor, nur fokussierten diese auf die Informationserhaltung, die Beweiswerterhaltung und Nachweisfähigkeit blieb außen vor. Zudem sind beide Zertifizierungsverfahren derzeit nicht regulatorisch determiniert.

Im Zuge der [EIDAS] wurden in Art. 34 und 40 [EIDAS] Anforderungen an die Bewahrung qualifizierter Signaturen und Siegel mittel qualifizierter Bewahrungsdienste definiert. Unter Bewahrungsdienste sind Vertrauensdienste zur beweiswerterhaltenden Aufbewahrung geschäftsrelevanter Aufzeichnungen zu verstehen. Bereitgestellt und betrieben werden diese durch (qualifizierte) Vertrauensdienste, die auf Basis der durch ETSI definierten Standards von nationalen Konformitätsbewertungs- bzw. Prüfstellen zertifiziert werden und ihre Dienste in der Folge im EWR anbieten können. Die Zertifizierung erfolgt national, die Anerkennung, wie auch bei den übrigen Vertrauensdiensten (z. B. Erzeugung qualifizierter elektronische Signaturen/Siegel/Zeitstempel oder Einschreib-/Zu-

stellendienste), international. Die Aufsicht obliegt in Deutschland der Bundesnetzagentur, als Prüfstellen fungieren aktuell der TÜVIT Informationstechnik GmbH, datenschutz Cert, SRC Security Research & Consulting GmbH und T-Systems International GmbH³ an der technischen Standardisierung ist das BSI umfangreich beteiligt. Die im Zuge von Mandat 460 bei ETSI entstehenden Standards für (qualifizierte) Bewahrungsdienste fokussieren vorrangig auf die Trust Service Provider, die wiederum entsprechend standardisierte Produkte für ihren Vertrauensdienst einzusetzen haben. Im Kern gehen auch die europäischen Standards von einer Aufbewahrung sowohl signierter als auch unsignierter Aufzeichnungen durch Bewahrungsdienste aus, deren Integrität und Authentizität langfristig durch kryptographische Signaturtechniken basierend auf elektronischen Signaturen, Siegeln oder qualifizierten Zeitstempeln oder technische Beweisdaten (Evidence Records gemäß [RFC4998], [RFC6283], [TR03125])) gesichert und erhalten wird. Die konkreten Verfahren variieren hierbei vom nicht nur in Deutschland angewendeten Hashbaumverfahren [Merk] [TR03125] bis hin zu aufwändigeren, einzelsignaturbezogenen Verfahren. Die nachfolgende Tabelle zeigt die Kerninhalte der für Bewahrungsdienste wesentlichen Standards im Überblick (*nächste Seite*):

3 Vgl. https://www.bundesnetzagentur.de/DE/Service-Funktionen/ElektronischeVertrauensdienste/Bewertungsstellen/KBS_nach_eIDAS/KBSnachEIDAS_node.html [Abruf 4. 10. 2019]

Standard	Geltungsbereich	Kerninhalte
ETSI EN 319 401	Alle Vertrauensdienste	<p>Grundlegende organisatorische und technische Sicherheitsanforderungen an Prozesse und Betrieb von Vertrauensdiensten</p> <p>Anforderungen hinsichtlich Betriebsorganisation, Business Continuity etc.</p> <p>Anforderungen an das interne Records Management des Vertrauensdienstes</p>
ETSI TS 119 511	Bewahrungsdienste	<p>Spezifische organisatorische und technische Anforderungen an Prozesse, eingesetzte Produkte und Betrieb von Bewahrungsdiensten</p> <p>Vertrauensdiensteanbieter muss die vollständige Kontrolle über alle Prozesse und Aufzeichnungen seines Bewahrungsdienstes besitzen, Bypässe, durch die Aufzeichnungen am Bewahrungsdienst vorbei, durch die Kunden abgelegt werden können, sind untersagt</p> <p>Ziel: Gewährleistung der langfristigen</p> <ul style="list-style-type: none"> • Prüfbarkeit kryptographischer Signatur-Techniken (Signaturen/Siegel/Zeitstempel/Evidence Records) • Bewahrung des Gültigkeitsstatus • Erhalt eines Proof of Existence der beweisrelevanten Daten sowie • Erhalt eines Proof of Existence aufbewahrten signierten oder unsignierten Daten <p>Speichermodelle:</p> <ul style="list-style-type: none"> • Mit Speicher (z. B. [TR03125]) • Mit temporärem Speicher • Ohne Speicher
ETSI TS 119 512	Bewahrungsdienste	<p>Definition einer generischen Referenzarchitektur für Bewahrungsdienste</p> <p>Definition Schnittstellen und Protokolle (die unmittelbar auf diejenigen der [TR03125] des BSI mappen)</p> <p>Integration des Hashbaumverfahren nach [RFC4998] [RFC6283] bzw. [TR03125] sowie der AIP-Formate XAIP nach [TR03125-F] resp. ASiC [ETSI EN319162]</p>

In die Standardisierung auf europäischer Ebene sind die fundierten Erfahrungen des BSI mit der [TR03125] in Deutschland unmittelbar eingeflossen. Im Zertifizierungsprozess sind vom möglichen (qualifizierten) Vertrauensdiensteanbietern sowohl die Vorgaben nach [EN319401] als auch [ETSI TS119511] zu erfüllen. Im Kontext [ETSI TS119511] sind demgemäß Produkte nach dem Stand der Technik, in Deutschland [TR03125], einzusetzen. Die Prüfung selbst erfolgt durch die nationale Konformitätsbewertungsstelle anhand dezidierter Prüfkataloge, welche die europäischen Standards weiter untersetzen. Diese nationalen Prüfkataloge befinden sich aktuell in Entwicklung und werden in naher Zukunft von der Bundesnetzagentur als nationaler Aufsichtsstelle veröffentlicht. Da sich die [ETSI TS119512] im Kern an das vom (qualifizierten) Bewahrungsdienst eingesetzte Produkt richtet, bildet dieser Standard die Brücke zu den bereits vorliegenden nationalen wie internationalen Standards im Kontext beweiswerterhaltender Langzeitspeicherung. Die aktuell in Entwicklung befindliche [TR03125] Version 1.2.2 und 1.3 nimmt die Maßgaben der [ETSI TS119512] unmittelbar auf und ermöglicht so eine [EIDAS] konforme beweissichere Langzeitspeicherung.

BSI TR-ESOR (Produkt)

Die [TR03125] des BSI, aktuell Version 1.2.1, gilt als verbindlicher Stand der Technik zur beweiswerterhaltenden Aufbewahrung in Deutschland. Sie beschreibt zum einen eine generische, modulare wie skalierbare Referenzarchitektur einer Middleware zur Beweiswerterhaltung, zum anderen die notwendigen verpflichtenden sowie optionalen Prozesse, Informationspakete und Schnittstellen sowie Formate für technische Beweisdaten (Evidence Records). Die Zertifizierung konkreter Produkte ermöglicht es dem Anwender, die Konformität von Marktlösungen gegenüber der TR transparent zu erkennen. Darüber hinaus werden vom BSI Testwerkzeuge entwickelt, die auf Basis der definierten Schnittstellen und Formate die technische Interoperabilität zwischen den Herstellern erleichtern [Ko13], [KoKuSc18], [Ko14], [KuSc16]. Die TR selbst beruht resp. integriert nationale wie internationale Standards, so insbesondere [ISO24727], [ISO14533], [RFC 4998] [RFC6283], [ISO13527] [ISO14721], [ETSI TS119511], [ETSI TS119512]. Die [TR03125] ist branchenübergreifend zur beweissicheren Langzeitspeicherung etabliert. Umfassende Anwendungsbeispiele finden sich u. a. im Public Sector, Gesundheitswesen, Luft- und Raumfahrt, Finanzindustrie oder dem Energiesektor. Aufbau und Dokumente sowie die Referenzarchitektur der [TR03125] zeigt folgende Grafik.

Abb. 5, Seite 231

Derzeit läuft die Weiterentwicklung der [TR03125] hinsichtlich der Bewahrungsdienste sowie zur optimierten Aufnahme großer Datenmengen (>200GB) und ermöglicht so u.a. weitere Archivdaten-Container. Die wesentlichen Anpassungen beschreibt die nachstehende Tabelle.

Version	Kerninhalt	Erscheinungsdatum (geplant)
v1.2.1	<p>Editorielle Anpassung auf [EIDAS] Änderungen an Funktionen des Krypto-Moduls zur Signaturprüfung und Einholung der beweisrelevanten Daten</p> <ul style="list-style-type: none"> • Schalenmodell und Kettenmodell • Anforderung qualifizierter Zeitstempel und optional elektronischer Signaturen/Siegel beim qualifizierten Vertrauensdiensteanbieter 	2018
v1.2.2	<p>AIP (große Datenmengen) Ergänzung eines logischen XAIP-Containers Ergänzung von ASiC-Container als Profilierung von ASiC-E [EN319162] Ergänzung der Referenzarchitektur bzgl. logischen AIP Präzisierung der zentralen Eingangsschnittstelle S.4 zur Verwendung von MTOM, Protokoll nach [ETSI TS119512] als weitere Eingangsschnittstelle neben der bestehenden S.4, auch auf Basis REST und JSON Bereitstellung des Testwerkzeugs zur Interoperabilität technischer Beweisdaten (Evidence Records)</p>	Ca. Q2/2019
v1.3	<p>Anpassung des Zertifizierungsschemas mit Fokus auf die technische Interoperabilität Open-Source-Modul zur Transformation der PreservationAPI [ETSI TS119512] auf S.4 Editorielle Anpassungen an den Dokumenten der TR</p>	Ca. Q1/2020

Zusammenwirken Bewahrungsdiensten und Produktstandards

Die Standardisierung hinsichtlich Bewahrungsdiensten sowie den von diesen eingesetzten Produkten und Systemen ist komplementär. Die europäischen Maßgaben seitens ETSI [EN319401] und [ETSI TS119511] setzen die organisatorischen, prozessualen sowie technischen Vorgaben an den (qualifizierten)

Vertrauensdienst, also den Service-Provider oder Betreiber. [ETSI TS119511] definiert zudem grundlegende Anforderungen ans eingesetzte System durch Verweis auf die [ETSI TS119512], die national durch die jeweiligen Systemstandards wie die [TR03125] in Deutschland integriert wird. Die nachstehende Abbildung zeigt das Zusammenwirken im Detail.

Abb. 6, Seite 232

Fazit und Ausblick

Sowohl die Regelungen der [EIDAS]-Verordnung als auch die hierauf basierenden ETSI-Normen schaffen die regulatorische wie technische Basis für einen digitalen Vertrauensraum im europäischen Binnenmarkt [BuDru18]. Untersetzt durch fachspezifische Regularien und Standardisierungen in Records Management und beweisicherer Langzeitspeicherung liegt ein rechtlich-organisatorisch wie valides fachlich-technisches Framework vor, um elektronische Prozesse Ende zu Ende, also Kunde-Behörde/Unternehmen-Kunde, sicher, nachweisbar und damit vertrauenswürdig bis zum Ablauf der geltenden Aufbewahrungsfristen zu realisieren.

Der aktuelle Schwerpunkt der Standardisierung im DIN und ISO liegt daneben insbesondere auf der Definition von Anforderungen an Records Management und Beweiswerterhalt in Blockchain. Ziel ist es, Kriterien zu definieren, um diese disruptive Technologie für vertrauenswürdige Transaktionen einsetzen zu können und den Langzeiterhalt der geschäftsrelevanten Aufzeichnungen dabei ebenso zu gewährleisten [KoBeScKu18], (DINSPEC31648).

Literatur

- | | |
|-----------------------|---|
| [2015/1506/EU] | DURCHFÜHRUNGSBESCHLUSS (EU) 2015/1506 DER KOMMISSION zur Festlegung von Spezifikationen für Formate fortgeschrittener elektronischer Signaturen und fortgeschrittener Siegel, die von öffentlichen Stellen gemäß Artikel 27 Absatz 5 und Artikel 37 Absatz 5 der Verordnung (EU) Nr. 910/2014 des Europäischen Parlaments und des Rates über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt anerkannt werden, 8. September 2015 |
| [Az. 6 K 691/14.WI.A] | VG Wiesbaden, 26.09.2014 - 6 K 691/14.WI.A |
| [Az. 6 K 808/17.WI.A] | VG Wiesbaden, 09.08.2017 - 6 K 808/17.WI.A |
| [BuDru18] | Vertrauensraum in der Digitalisierung. Bundesdruckerei (Hrsg.). Berlin 2018 |
| [BSIGS] | BSI-Grundschutzkompendium. Bundesamt für Sicherheit in der Informationstechnik. Bonn 2018 |

[DIN31644]	DIN 31644:2012 Information und Dokumentation – Kriterien für vertrauenswürdige digitale Langzeitarchive. 2012
[DIN31647]	DIN 31647:2015 Beweiswerterhalt kryptografisch signierter Dokumente, 2015.
[DINSPEC31648]	DIN SPEC 31648: Kriterien für vertrauenswürdige Transaktionen - Records Management und Beweiswerterhaltung in Blockchain
[EUDSGVO]	VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)
[EGOVG]	Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz - EGovG) vom 25.07.2013
[EASA]	Easy Access Rules for Airworthiness and Environmental Certification. (Regulation (EU) No 748/2012). Part 21.A.55. published February 2018.
[EIDAS]	VERORDNUNG (EU) Nr. 910/2014 DES EUROPÄISCHEN PARLAMENTS UND DES RATES über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG“ vom 23.07.2014
[eJusticG]	Gesetz zur Förderung des elektronischen Rechtsverkehrs mit den Gerichten vom 10. Oktober 2013
[EN319122]	ETSI EN 319 122 – {1,2,3}, Electronic Signatures and Infrastructures (ESI); CAdES digital signatures, ETSI V1.1.1, (2016-04)
[EN319132]	ETSI EN 319 132 – {1,2}, Electronic Signatures and Infrastructures (ESI); XAdES digital signatures, ETSI V1.1.1, (2016-04)
[EN319142]	ETSI EN 319 142 – {1,2}, Electronic Signatures and Infrastructures (ESI); PAdES digital Signatures, ETSI V1.1.1 (2016-04)
[EN319162]	ETSI EN 319 162 – {1,2}, Electronic Signatures and Infrastructures (ESI); Associated Signature Containers (ASiC), ETSI V1.1.1 (2016-04)

- [EN319401] ETSI EN 319 401 - General Policy Requirements for Trust Service Providers. ETSI v2.2.1 (2018-04)
- [EU-DLR] RICHTLINIE 2006/123/EG DES EUROPÄISCHEN PARLAMENTS UND DES RATES über Dienstleistungen im Binnenmarkt“ vom 12.12.2006 [EUDSGVO]
VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)
- [EuroSOX] RICHTLINIE 2006/43/EG DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 17. Mai 2006 über Abschlussprüfungen von Jahresabschlüssen und konsolidierten Abschlüssen, zur Änderung der Richtlinien 78/660/EWG und 83/349/EWG des Rates und zur Aufhebung der Richtlinie 84/253/EWG des Rates
- [FDA] CFR - Code of Federal Regulations Title 21. ITLE 21--FOOD AND DRUGS CHAPTER I--FOOD AND DRUG ADMINISTRATION DEPARTMENT OF HEALTH AND HUMAN SERVICES SUBCHAPTER A--GENERAL PART 11 ELECTRONIC RECORDS; ELECTRONIC SIGNATURE
- [Fi06] S. Fischer-Dieskau: Das elektronisch signierte Dokument als Mittel zur Beweissicherung, Baden-Baden, 2006.
- [GXP] Sammlung von Guten Arbeitspraxis Richtlinien, z. B. GCP – Good Clinical Practice oder GLP – Good Laboratory Practice etc
- [He18] Henne, Thomas: Juristische Anforderungen an die Beweiswerterhaltung bei digitaler Archivierung. 23. Archivwissenschaftliches Kolloquium. Marburg 2018
- [ISO13527] ISO 13527:2010, Space data and information transfer systems -- XML formatted data unit (XFDU) structure and construction rules, 2010
- [ISO14533-1] ISO 14533-1:2014, Processes, data elements and documents in commerce, industry and administration -- Long term signature profiles -- Part 1: Long term signature profiles for CMS Advanced Electronic Signatures (CAeS)

- [ISO14533-2] ISO 14533-2:2012 Processes, data elements and documents in commerce, industry and administration -- Long term signature profiles -- Part 2: Long term signature profiles for XML Advanced Electronic Signatures (XAdES)
- [ISO14533-3] ISO 14533-3:2017 Processes, data elements and documents in commerce, industry and administration -- Long term signature profiles -- Part 3: Long term signature profiles for PDF Advanced Electronic Signatures (PAdES)
- [ISO14721] ISO 14721:2012, Space data and information transfer systems – Open archival information system – Reference model, 2nd Edition, 2012
- [ISO15489] ISO 15489-1:2016: Information and documentation - Records management - Part 1: Concepts and principles. 2016
- [ISO30300] ISO 30300:2011. Information and documentation - Management systems for records - Fundamentals and vocabulary
- [ISO30301] ISO 30301:2011, Information and documentation - Management systems for records - Requirements. 2011
- [ISO30302] ISO 30302:2015, Information and documentation -- Management systems for records -- Guidelines for implementation
- [Ko13] U. Korte, S. Schwalm, D. Hühnlein: Vertrauenswürdige und beweiswerterhaltende Langzeitspeicherung auf Basis von DIN 31647 und BSI TR-03125, Informatik 2013, GI-LNI, P220, ISBN 978-3-88579-614-5, S. 550-566, 2013
- [KoScDH14] U. Korte, S. Schwalm, D. Hühnlein: Standards for the preservation of evidence and trust. Proceedings Archiving 2014, Springfield 2014, S. 9-14
- [Ko14] U. Korte, S. Schwalm, D. Hühnlein: Standards und Lösungen zur langfristigen Beweiswerterhaltung, DACH-Security 2014, S. 46-58, Frechen 2014
- [Ko16] U. Korte. S.Schwalm. T. Kusber. D. Hühnlein: Beweiswerterhaltung im Kontext eIDAS - eine Case Study. DACH-Security 2016, Frechen 2016 S. 379-392

- [Ko17] U. Korte, S. Schwalm, T. Kusber, D. Hühnlein, M. Prechtl, B. Wild: Datenpakete zur Informations- und Beweiserhaltung. Ein Vergleich. DACH-Security 2017. Frechen 2017 S. 291-303
- [KSDV15] T. Kusber, S. Schwalm, A. Dörner, T. Vogt, Die Bedeutung der eIDAS-Verordnung für Unternehmen und Behörden. Neue Chancen und Herausforderungen für vertrauenswürdige elektronische Geschäftsprozesse in Europa, Berlin, 2015
- [KuSc16] T. Kusber, S. Schwalm: Elektronische Langzeitspeicherung als SOA-Dienst – Kernelement eines vertrauenswürdigen Informationsmanagements. INFORMATIK 2016 S. 869-882
- [KoScKu18] U. Korte, S. Schwalm, T. Kusber: Vertrauenswürdige E-Government – Anforderungen und Lösungen zur beweiserhaltenden Langzeitspeicherung. 23. Archivwissenschaftliches Kolloquium. Marburg 2018
- [KoBeScKu18] U. Korte, Christian Berghoff, Steffen Schwalm, Tomasz Kusber: Langfristige Beweiserhaltung und Datenschutz in der Blockchain, DACH-Security 2018, S. 177-192, Gelsenkirchen, 2018
- [LOTAR] <http://lotar-international.org/home.html>
- [Merk] R. Merkle: Protocols for Public Key Cryptosystems, Proceedings of the 1980 IEEE Symposium on Security and Privacy (Oakland, CA, USA), SS. 122-134, 1980
- [MiniKGovG] Minikommentar zum Gesetz zur Förderung der elektronischen Verwaltung sowie zur Änderung weiterer Vorschriften. Bundesministerium des Innern. Berlin 2013
- [OZG] Onlinezugangsgesetz vom 14. August 2017 (BGBl. I S. 3122, 3138)
- [PSD2] RICHTLINIE (EU) 2015/2366 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 25. November 2015 über Zahlungsdienste im Binnenmarkt, zur Änderung der Richtlinien 2002/65/EG, 2009/110/EG und 2013/36/EU und der Verordnung (EU) Nr. 1093/2010 sowie zur Aufhebung der Richtlinie 2007/64/EG

- [RFC4998] T. Gondrom, R. Brandner, U. Pordesch: Evidence Record Syntax (ERS), IETF RFC 4998, <http://www.ietf.org/rfc/rfc4998.txt>, August 2007.
- [RFC6283] A. J. Blazic, S. Saljic, T. Gondrom: Extensible Markup Language Evidence Record Syntax (XMLERS), IETF RFC 6283, <http://www.ietf.org/rfc/rfc6283.txt>, Juli 2011
- [Ro07] A. Rossnagel: Langfristige Aufbewahrung elektronischer Dokumente, Anforderungen und Trends, Baden-Baden, 2007
- [SR019510] ETSI SR 019 510, Electronic Signatures and Infrastructures (ESI); Scoping Study and framework for standardization of long-term data preservation services, including preservation of/with digital signatures, ETSI V1.1.1 (2017-05)
- [To07] Peter M. Toebak: Records Management. Ein Handbuch. Baden 2007
- [TR-03125] BSI: Beweiswerterhaltung kryptographisch signierter Dokumente (TR-ESOR), TR 03125, <https://www.bsi.bund.de/tr-esor>.
- [TR-03138] BSI Technische Richtlinie 03138 Ersetzendes Scannen. Version 1.2, 2018
- [ETSI TS119511] Policy and security requirements for trust service providers providing long-term preservation of digital signatures or general data using digital signature techniques
- [ETSI TS119512] Protocols for trust service providers providing long-term data preservation services
- [VDG] Vertrauensdienstegesetz vom 18. Juli 2017 (BGBl. I S. 2745), das durch Artikel 2 des Gesetzes vom 18. Juli 2017 (BGBl. I S. 2745) geändert worden ist
- [VS18] T. Vogt, S. Schwalm: eIDAS-ready?. Status Quo und Roadmap zur Umsetzung elektronischer Vertrauensdienste in Behörden und Unternehmen. Berlin 2018
- [We18] M. Weber, T Vogt, W. Krogel, S. Schwalm: Records Management nach ISO 15489. Einführung und Anleitung. Berlin 2018

[WeSc18]

M. Weber, S. Schwalm: Records Management und die Internationale Norm ISO 15489:2016. DIN-Normenausschuss Information und Dokumentation (NID). Regelkonforme Erzeugung, Speicherung und Nutzung digitaler Aufzeichnungen. in: DIN-Mitteilungen, Heft 9/2018, S. 10-13

[ZPO]

Zivilprozessordnung in der Fassung der Bekanntmachung vom 5. Dezember 2005 (BGBl. I S. 3202; 2006 I S. 431; 2007 I S. 1781), die zuletzt durch Artikel 11 Absatz 15 des Gesetzes vom 18. Juli 2017 (BGBl. I S. 2745) geändert worden ist

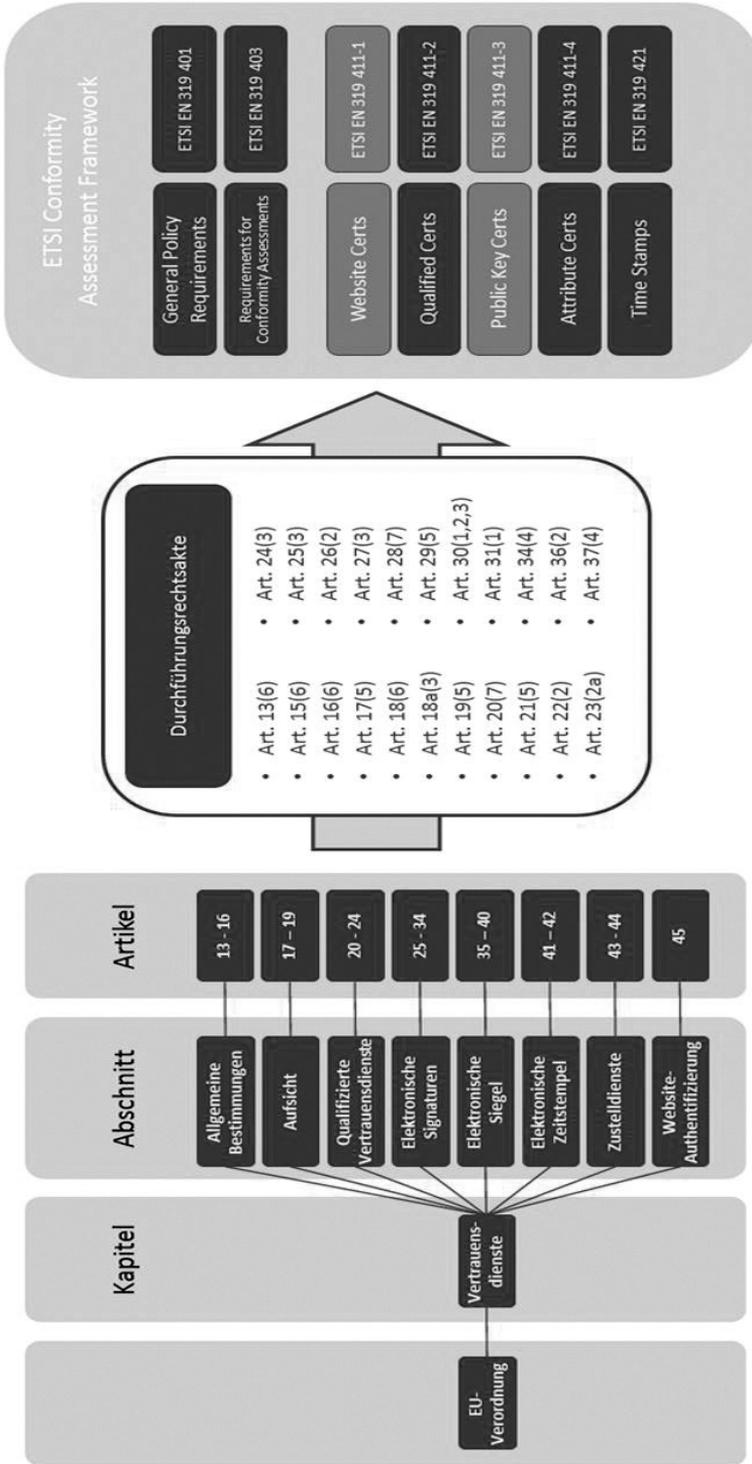


Abb. 1: Beziehung zwischen eIDAS-Verordnung, Implementing Acts und ETSI-Standardisierung

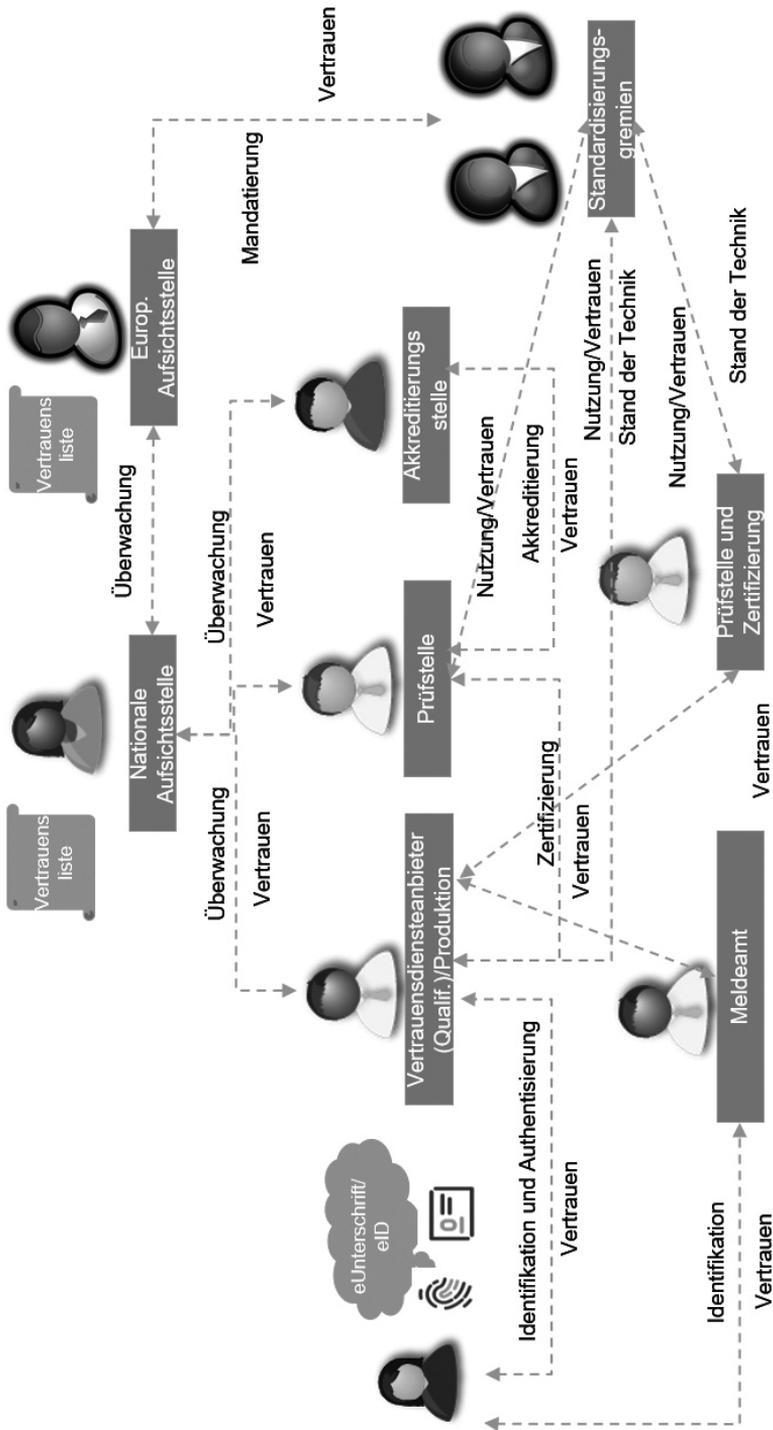


Abb. 2: Vertrauenswürdigkeit geschäftsrelevanter Transaktionen und Aufzeichnungen am Beispiel einer Vertrauenskette gem. [EIDAS]

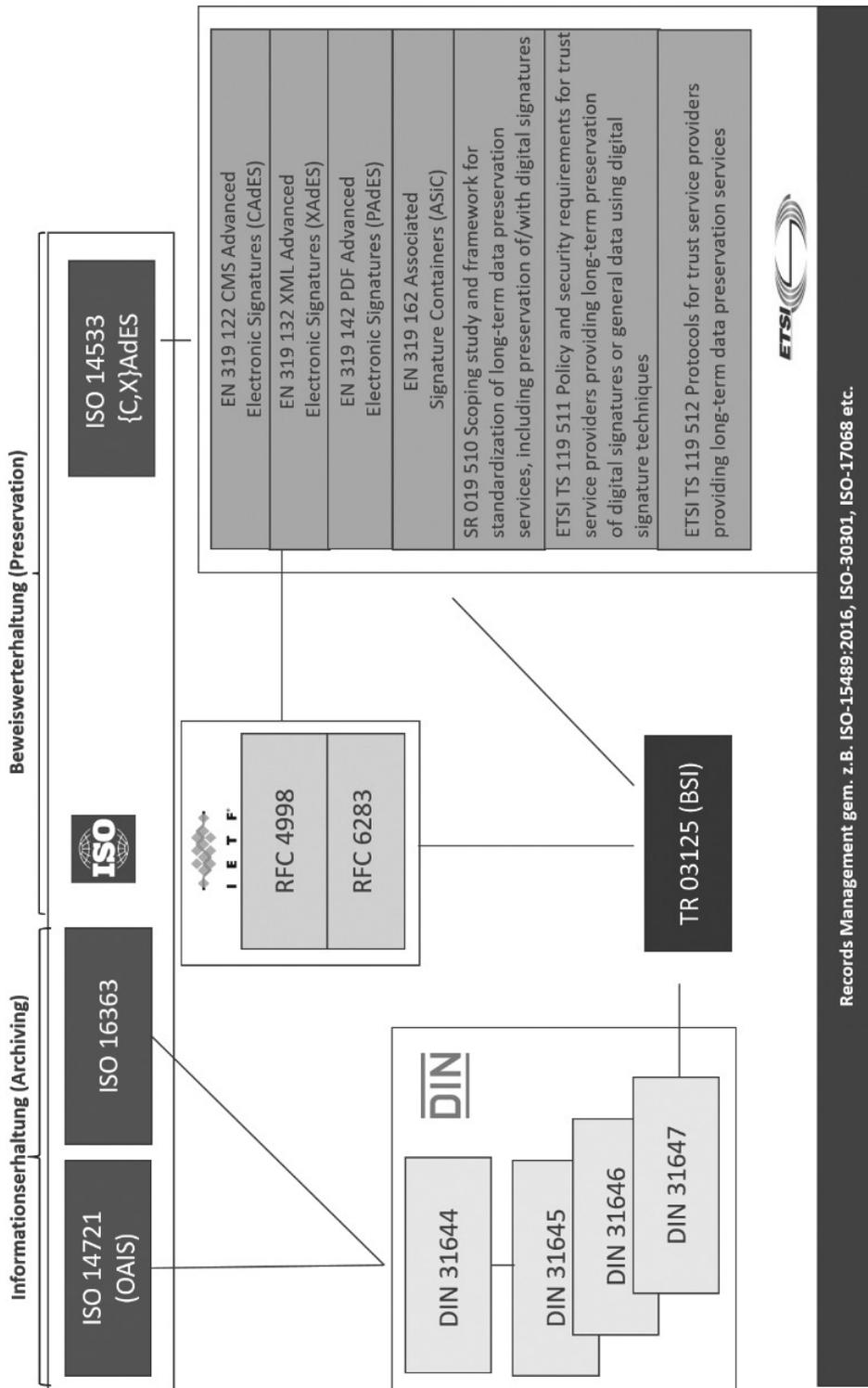


Abb. 3: Überblick wesentlicher Standards und Normen zum langfristigen Nachweis digitaler Transaktionen und Aufzeichnungen

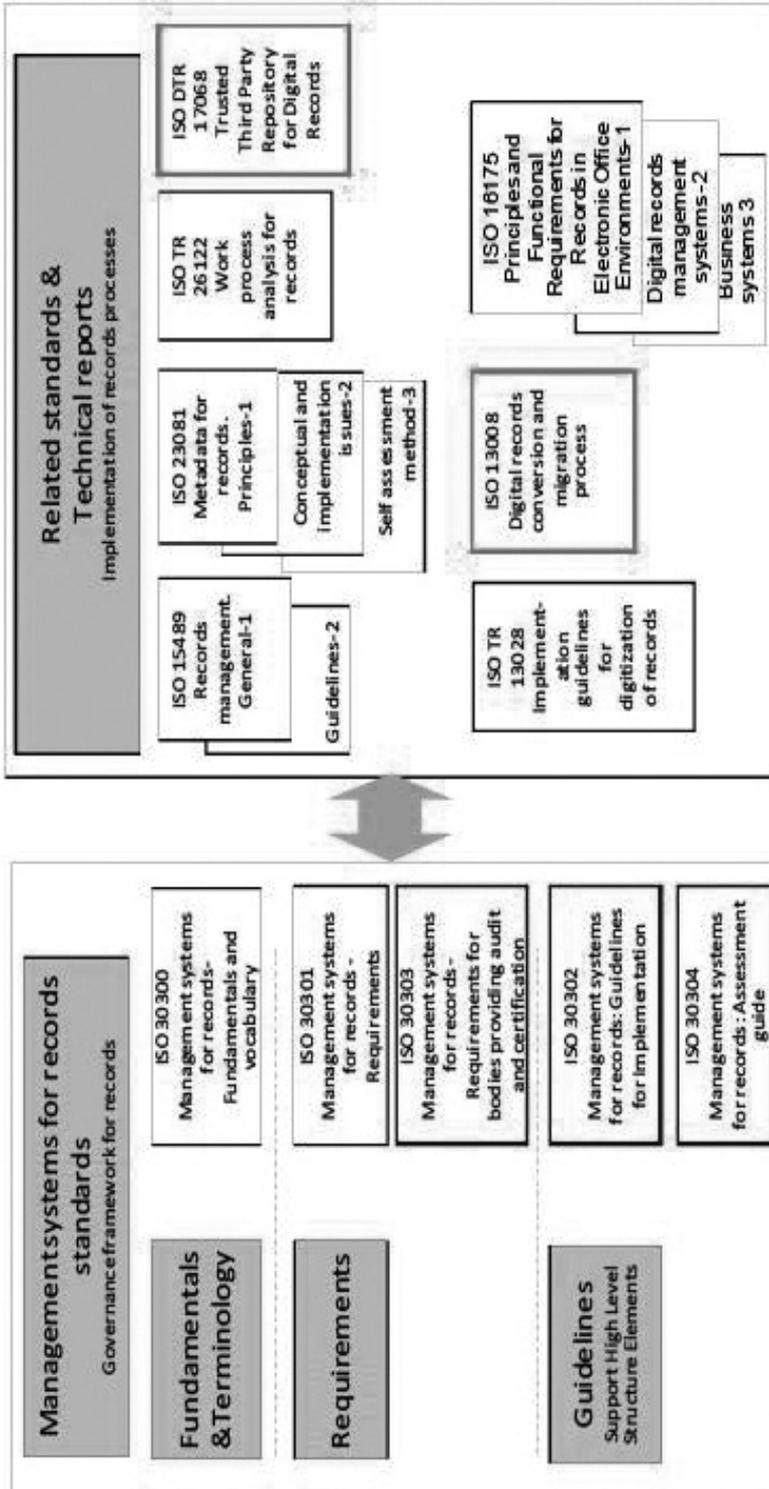
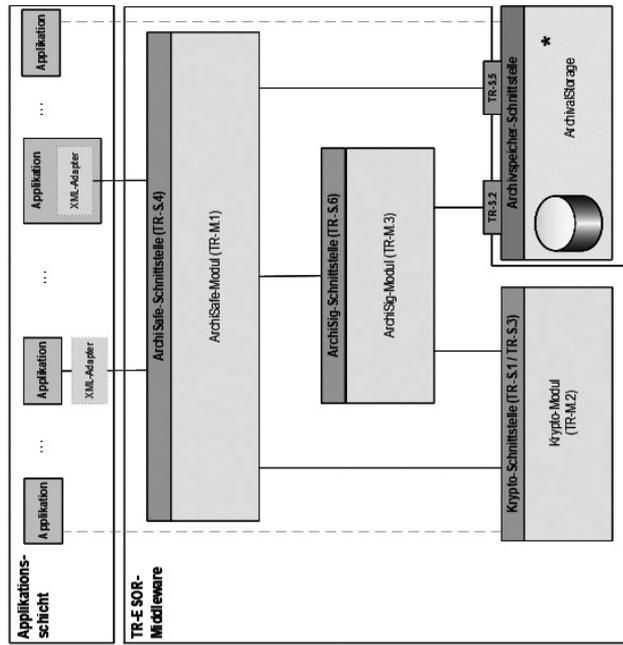
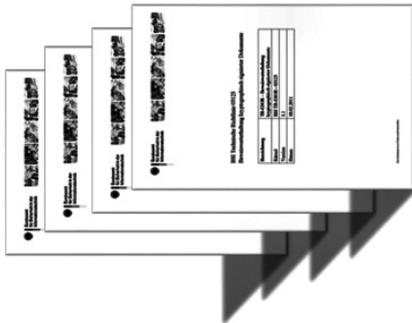


Abb. 4: ISO-Standardisierungsframework zum Records Management



- TR-ESOR Hauptdokument
- TR-ESOR-M.1 ArchiSafe Modul
 - TR-ESOR-M.2 Krypto Modul
 - TR-ESOR-M.3 ArchiSig Modul
 - TR-ESOR-S Schnittstellen
 - TR-ESOR-B Bundesbehördenprofil
 - TR-ESOR-F Formate
 - TR-ESOR-E Konkretisierung d. Schnittstellen auf Basis des eCard-API Frameworks
 - TR-ESOR-VR Verifikationsreport für ausgewählte Datenstrukturen
 - TR-ESOR-ERS Profilierung der Evidence Records auf Basis von RFC 4998 und RFC 6283
 - TR-ESOR-XBDP Profilierung des XAIP mit XBARCH, XDOMEA und PREMIS
 - TR-ESOR-C.1 Testspezifikation „Funktionale Konformität“
 - TR-ESOR-C.2 Testspezifikation „Technische Konformität“
 - TR-ESOR-C.3 Testspezifikation „Bundesbehörden-Profil“



**Aktuelle Version: 1.2.1
(v1.3 in Entwicklung)**

Abb. 5: Dokumente und Referenzarchitektur BSI TR-3125 TR-ESOR

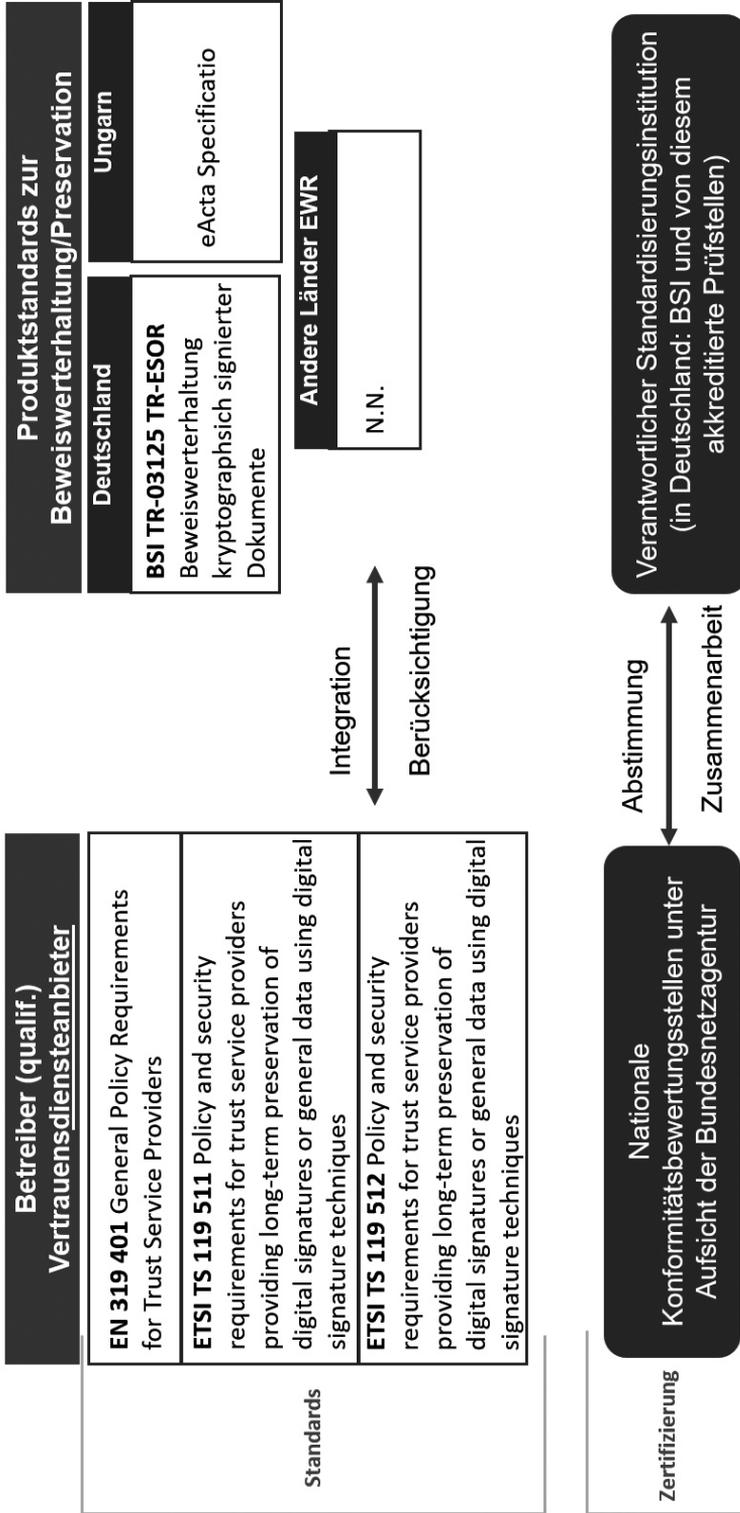


Abb. 6: Zusammenhang zwischen Zertifizierung von Bewahrungsdiensten und deren Produkten

DIE NESTOR-AG „KOOOPERATION DER ARCHIVE“ UND DER NESTOR-ARCHIVSTANDARD

Christine Gigler

Die nestor-AG „Kooperation der Archive“ Intention und Ergebnis

Im Januar 2016 konstituierte sich die nestor-Arbeitsgruppe „Kooperation der Archive“¹ unter der Leitung Christian Keitels (Landesarchiv Baden-Württemberg) mit der Zielsetzung, klassische Archive bei der konkreten Verwirklichung von Kooperationen zu unterstützen.² In den beiden darauf folgenden Jahren wurden in drei Unterarbeitsgruppen verschiedene Themen bearbeitet, die sich aus der Anfangsdiskussion der diversen Zielvorstellungen ergeben hatten. Jede Unterarbeitsgruppe entwickelte ein Produkt, das dazu beitragen soll, klassische Archive bei der Vernetzung im Hinblick auf die digitale Archivierung zu stärken.

Die Resultate wurden im Februar 2018 in zwei Bänden „nestor-materialien“ und auf der Website von nestor veröffentlicht. Der nestor-materialien-Band 21 „Gemeinsam handeln – Vorschläge für Archive“³ beschreibt schon existierende Kooperationsmodelle und analysiert anhand von sechs Fallbeispielen Erfolgsfaktoren für eine gelungene Zusammenarbeit, aber auch Problemfelder, die eine solche erschweren können. Der Band bietet außerdem Starthilfen für die digitale Archivierung in Form von FAQs, die an unterschiedlichen Erfahrungslevels Einstiegspunkte in die Thematik ermöglichen und von „Ich habe keine Ahnung, wie fange ich an?“ bis hin zur Beantwortung der Frage, wo man passende Softwaretools finden kann, reichen. Nestor-materialien 22⁴ stellt den nestor-Standard für Archive vor, auf den im Folgenden näher eingegangen wird. Die erwähnte kommentierte Linksammlung für Neueinsteiger⁵ und ein Kontakt-

1 <https://wiki.dnb.de/display/NESTOR/AG+Kooperation+der+Archive> [Abruf 27. 5. 2019].

2 Die Gründung der Arbeitsgruppe ging auf eine Anregung des nestor-Workshops „nestor und die Vernetzung der Archivsparte: Sachstand, Wünsche und Perspektiven“, der am 16. April 2015 in Koblenz stattgefunden hatte, zurück, siehe <http://www.dnb.de/Subsites/nestor/DE/Veranstaltungen/TermineNestor/vernetzungArchivsparte.html> [Abruf 27. 5. 2019].

3 Gemeinsam handeln – Vorschläge für Archive, verfasst und herausgegeben von der nestor-Arbeitsgruppe Kooperation der Archive (nestor-Materialien 21). [Frankfurt/Main] 2018, <http://nbn-resolving.de/urn:nbn:de:0008-2018020837> [Abruf 27. 5. 2019].

4 Gemeinsam handeln – nestor-Standard für Archive, verfasst und herausgegeben von der nestor-Arbeitsgruppe Kooperation der Archive (nestor-Materialien 22). [Frankfurt/Main] 2018, <http://nbn-resolving.de/urn:nbn:de:0008-2018020847> [Abruf 27. 5. 2019].

5 <https://wiki.dnb.de/pages/viewpage.action?pageId=133368003> [Abruf 27. 5. 2019].

formular für Kooperationsprojekte⁶ sind unter dem Button „Services“ auf der Startseite von nestor zu finden und ergänzen die beiden nestor-materialien.

Hinsichtlich ihres Umsetzungspotenzials positionieren sich die beiden nestor-materialien-Bände einerseits zwischen den eher abstrakten Grundsatzüberlegungen der nestor-Kriterienkataloge und andererseits den Bedürfnissen nach praktischen Lösungen einzelner Archiveinrichtungen.

Der nestor-Archivstandard

Bevor der neu zu etablierende nestor-Archivstandard vorgestellt wird, soll auf die Frage eingegangen werden, warum es notwendig erscheint, weitere Standards einzuführen, und an wen sich das hier präsentierte Verfahren wendet.



Abb. 1: Cover nestor-Materialien Band 22

Gerade auf dem Gebiet der digitalen Archivierung besteht nach wie vor Abstimmungs- und Koordinationsbedarf. Die Probleme in diesem Bereich betreffen meist nicht nur ein einzelnes Archiv, sondern in der Regel jeweils zahlreiche andere. Es herrscht weitgehend Konsens darüber, dass Kooperationen – vor allem

6 https://www.langzeitarchivierung.de/Webs/nestor/DE/Services/services_node.html;jsessionid=994A801F6C8272156EBD70BF475D4A3B.internet572 [Abruf 13. 11. 2019].

für mittlere und kleinere Archiveinrichtungen, aber nicht nur für diese – dabei der Schlüssel zum Erfolg sind.⁷ Genau hier setzt der nestor-Archivstandard an: Er gibt Archiven die Möglichkeit, ihre Interessen gebündelt zu artikulieren. Der von der nestor-Arbeitsgruppe „Kooperation der Archive“ erstellte Leitfaden bietet nämlich ein Verfahren, anhand dessen klassische Archive ihre konkreten Anforderungen auf dem Gebiet der digitalen Archivierung formulieren und gemeinsam nach außen hin vertreten können. Am Ende eines solchen Verfahrens steht sodann als konsolidierte Forderung der (deutschen) Archivcommunity ein nestor-Archivstandard. Das Verfahren richtet sich also zunächst an die sogenannten „klassischen“ Archive und nicht an andere Gedächtnisinstitutionen, und es versteht sich als Ergänzung bestehender Abstimmungsverfahren. Diese und Exponenten anderer Bereiche wie etwa der Wirtschaft können sich zu einem späteren Zeitpunkt mittels Stellungnahmen beteiligen.

Die Adressanten eines solchen Archivstandards sind etwa Archivträger im Hinblick auf die Anforderungen digitaler Archive, IT-Kooperationspartner bezüglich der Ansprüche von Archiven bei Ausschreibungen und die Hersteller von Software, wenn es z. B. um spezielle Schnittstellen geht.

Prozessschritte

Wie sieht das Verfahren zur Erlangung eines solchen nestor-Archivstandards aus?

Es kann von einem klassischen Archiv allein oder mehreren Archiven veranlasst werden, indem ein schriftlicher Antrag an die zuständige nestor-AG Archivstandards gerichtet wird. Daraufhin bestimmt die nestor-AG Archivstandards aus ihrem Kreis einen Verfahrensverantwortlichen, der den Antrag prüft, und zwar unter den Gesichtspunkten Relevanz und Nachhaltigkeit des Anliegens, Klarheit der Zielvorstellungen und Stringenz der Argumentation. Er teilt seine Empfehlung den anderen Mitglieder der AG nestor-Archivstandards mit und alle treffen gemeinsam eine Mehrheitsentscheidung. Diese kann aus einer Ablehnung, einer Rückgabe des Antrags an den Initiator zwecks Überarbeitung oder der Annahme des Antrags bestehen. Gegen die Ablehnung kann der Antragsteller Einspruch erheben, worüber der/die Vorsitzende der AG mit seinem/ihrer Stellvertreter bestimmt.

.....

7 Siehe etwa Keitel, Christian: Warum ist Kooperation bei der digitalen Archivierung unumgänglich?. In: Ders.; Naumann, Kai (Hg.): Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“ (Werkhefte der staatlichen Archivverwaltung Baden-Württemberg A/24). Stuttgart 2013, S. 281–288, auch unter https://www.landesarchiv-bw.de/sixcms/media.php/120/59134/Keitel_Warum_ist_Kooperation_bei_der_digitalen_Archivierung_unumgaenglich.pdf [Abruf 27.5.2019]. Worm, Peter: Erfahrungen des LWL-Archivamts bei der Koordination von Projekten zum Langzeiterhalte elektronischer Daten, in: ebd., S. 297–309, auch unter https://www.landesarchiv-bw.de/sixcms/media.php/120/59138/Worm_Erfahrungen_des_LWL_Archivamts.pdf [Abruf 27. 5. 2019]; Büchler, Georg: So funktioniert archivische Kollaboration – und so nicht: In: Maissen, Anna Pia; Müller, Peter: Vernetzung und Kollaboration von Archiven. Vorträge des 75. Südwestdeutschen Archivtags am 18. und 19. Juni 2015 in Rotenburg am Neckar. Stuttgart 2016, S. 39–43, auch unter <https://kost-ceco.ch/cms/index.php?id=344,655,0,0,1,0> [Abruf 27. 5. 2019].

Wird der Antrag angenommen, dann wird der Initiator vom Verfahrensverantwortlichen aufgefordert, eine Entwurfsgruppe zur Ausarbeitung des Standards zusammenzustellen. Der Initiator hat auch bis zum Ende des Verfahrens die Federführung inne. Die Entwurfsgruppe sollte aus mindestens drei Mitgliedern aus unterschiedlichen Archiven bestehen; maximal sollte sie 15 Beteiligte zählen. Der Initiator, der die Entwurfsgruppe leitet, moderiert auch die Sitzungen, von denen es zumindest eine konstituierende und eine abschließende geben soll. Wenn es der Entwurfsgruppe erforderlich erscheint, kann sie Fachexperten konsultieren. Das kann in Form von schriftlichen oder mündlichen Expertisen sein oder indem diese zu Vorträgen eingeladen werden, damit die Mitglieder der Entwurfsgruppe diverse Kompetenzen ausbauen können. Die Meinung der Experten ist für die Entwurfsgruppe aber unverbindlich. Der fertige Entwurf, der im Konsens erarbeitet werden sollte, wird mit dem Verfahrensverantwortlichen abgestimmt, wobei nicht mehr als zwei Jahre seit der Konstituierung der Entwurfsgruppe vergangen sein sollten. Dieser kann die Entwurfsgruppe eventuell beauftragen, den Entwurf noch einmal zu überarbeiten. Können sich die beiden – also Verfahrensverantwortlicher und Entwurfsgruppe – nicht einigen, dann trifft die AG-nestor Archivstandards eine abschließende Entscheidung, wobei eine einfache Mehrheit ausreicht.

Der abgestimmte oder überarbeitete Entwurf wird jetzt wiederum vom Verfahrensverantwortlichen zur Kommentierung auf der nestor-Website publiziert. An dieser Stelle sind nicht nur die klassischen Archive, sondern alle Interessierten – Vertreter anderer Gedächtniseinrichtungen, Softwarehersteller, Mitarbeiter von Rechenzentren etc. – aufgerufen, Stellung zu nehmen. Gleichzeitig kann die Entwurfsgruppe zu einem Workshop einladen, um den Entwurf auch in diesem Rahmen zu diskutieren.

Nach Ablauf einer Frist von drei Monaten werden die eingelangten Kommentare von der Entwurfsgruppe erörtert. Falls sie Änderungsvorschläge nicht berücksichtigen möchte und strittige Punkte mit Kritikern nicht anderweitig klären kann, muss sie mit diesen eine Sitzung abhalten. Auf dieser Einspruchsitzung soll versucht werden, eine einvernehmliche Lösung herzustellen. Wobei die Entwurfsgruppe nicht verpflichtet ist, die kontroversen Aspekte zu akzeptieren.

Am Ende des Prozesses wird das Verfahren entweder eingestellt oder der finale Entwurf von der nestor-Geschäftsstelle als nestor-Archivstandard auf ihrer Internetseite veröffentlicht.

Die Frage, wie repräsentativ das Resultat dieses Verfahrens ist und welche Verbindlichkeit ein derartiger nestor-Archivstandard für sich beanspruchen kann, hat die Unterarbeitsgruppe „Kooperation der Archive“ in der Anfangsphase intensiv beschäftigt. Zuerst wurden verschiedene Varianten von Zustimmungsverfahren erwogen und diskutiert, wie sich ein Wählerkreis zusammensetzen sollte, ob Archive darüber abstimmen sollten, Verbände oder Einzelpersonen oder eine Mischung aus allen dreien. Dabei ging es um die Frage der Berechtigung und der Betroffenheit, welches Quorum gegeben sein müsste und wie ein solcher Abstimmungsprozess durchgeführt werden könnte. Diese Idee wurde von der Gruppe aber schließlich verworfen, denn sie gelangte zu der Überzeugung, dass die Repräsentativität durch die Mindestgröße der Entwurfsgruppe und die Beteiligungen in der Kommentierungsphase ausreichend sichergestellt ist.

Ausblick

Um das geschilderte Prozedere realisieren zu können, war schon von Beginn an die Einrichtung einer nachfolgenden Arbeitsgruppe nestor-Archivstandards vorgesehen. Diese neue Arbeitsgruppe hat sich mittlerweile konstituiert. Am 28. Mai 2019 fand bei der Deutschen Nationalbibliothek in Frankfurt am Main, also bei der nestor-Geschäftsstelle, die erste Sitzung statt. In der Folge wird es zeitnah zu einem Workshop und zu einem ersten Standardisierungsverfahren kommen.

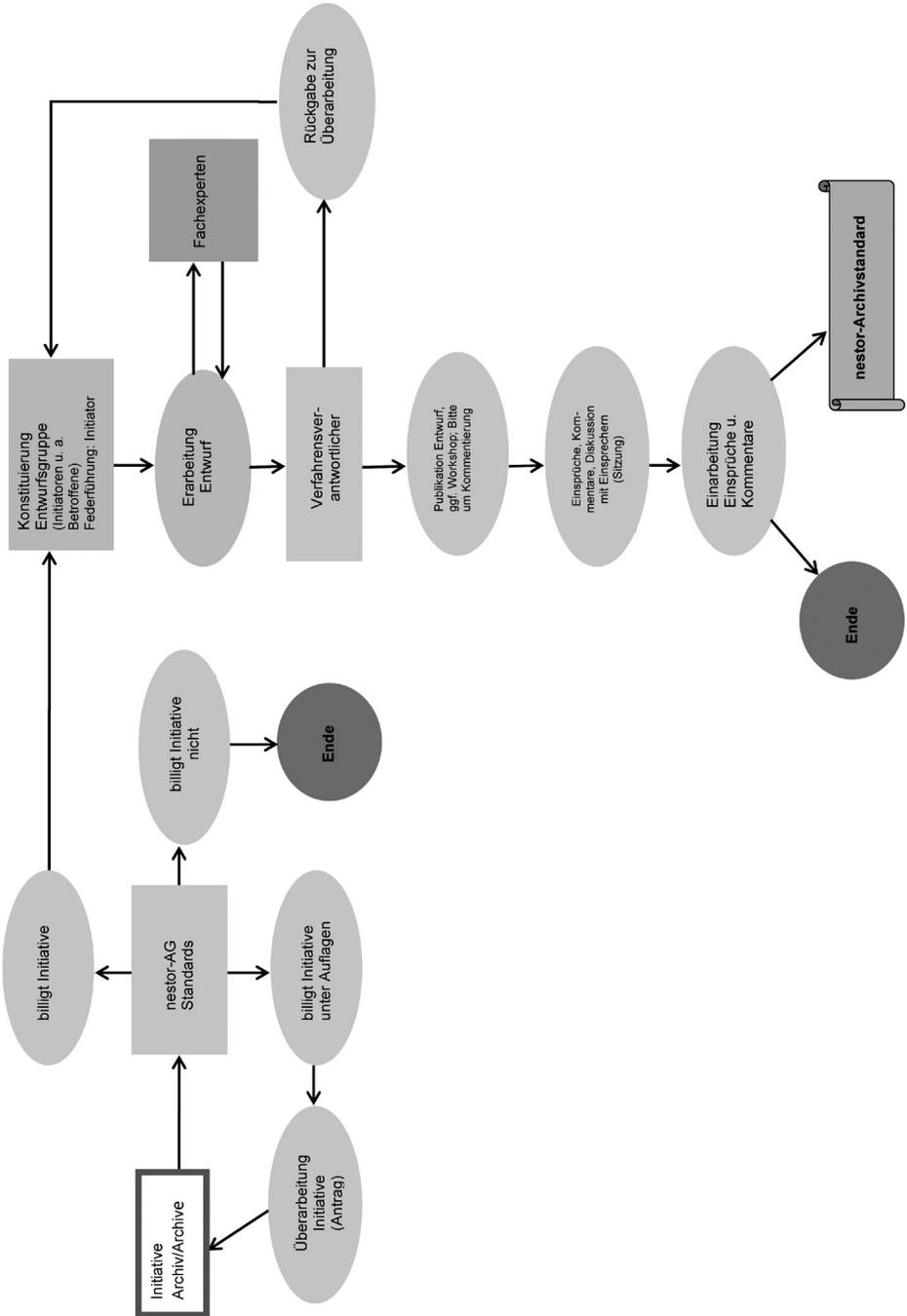


Abb. 2: Prozessdiagramm (nestor-AG Kooperation der Archive, Unter-AG 2)

AUDS-CAMP

DIGITAL OBJECTS

AND THEIR

PRESERVATION IN

ACADEMIC WORLD

DIGITAL OBJECTS AND THEIR PRESERVATION IN ACADEMIC WORLD

Summary

Zdeněk Vašek

The main theme of this Camp was the issue of long-term preservation of digital documents in the Academic World. Presented presentations and speeches of other participants focused on all aspects related to these documents. At the beginning of Camp were presented two presentations in which the authors described some current issues related to the topic of digital documents and ways of solving them.

Eliška Pavlásková from the Institute of the History and Archive of Charles University presented a paper entitled Digital Documents at Charles University. She described the current issues and challenges associated with digital documents at the university. In particular, she presented various types of digital documents and the possibilities for long-term preservation for each group of these documents. She also dealt with the management of digital documents and the problems associated with the fragmentation of academic world and fragmentation of originators. Charles University is the largest university in Czech Republic. There are currently 17 faculties at the University, plus 14 other departments. As such the University is very heterogenous institution and its complexity has serious consequences for management and preservation of digital materials of all kinds. There is not only need for different preservation strategies for different types of material, but also a need for careful coordination between departments of the University.

The second contribution was presented by Martin Lhoták from the Library of the Academy of Sciences of the Czech Republic. The library is an independent academic institution that deals with the administration of library services as well as management of digital document and research of this topic. Martin Lhoták presented the services that the library provides to other institutes of the Academy of Sciences as well as a research project in which, together with other Czech institutions, it develops the ARCLib open source LTP tool for memory institutions.¹ He introduced the basic functionality of this LTP system. The Library of the Academy of Sciences realizes extensive digitization of the library collection and also provides data processing from other institutes of the Academy of Sciences. The library must respect national standards for the digitization of books, but also manage scientific data that is very different.

.....
1 Project Page <https://github.com/LIBCAS/ARCLib/wiki> [Accessed 18. 6. 2019].

Both speakers agreed that the academic world is very specific to digital document management. The academic world and its documents are in many ways different from typical documents that are managed by archives and libraries. The documents include both the results of the normal day-to-day administrative agenda of the institutions and the scientific data and documents that are used for teaching or as a qualification work. The basic feature of the academic world is the considerable fragmentation and thus the large number of different originators who differ in their needs. Data requires different strategies for long-term preservation. Curators need to know their characteristics and usage to choose the right strategy for each type of data. The faculties and academic institutes have independent policies and should be taken into account when planning long-term preservation practices.

The discussions asked the two referring, in particular, to the practical aspects of managing and preserving digital documents in their institutions. They were interested in long-term preservation strategies. The participants agreed that the fragmentation of academic world is a serious problem and highlighted the need for a strong mandate for digital document management institutions. They pointed out in particular the need for institutional policy in the field of long-term preservation. In conclusion, both keynote speakers and participants stated that processes have been set up in these institutions so that the preconditions for successful application of recommendations for long-term protection of digital documents can be fulfilled. At the same time, they noted that digital documents that arise in the academic world require a special approach that is similar to that of classical archives and libraries, but has its own specificities.

AUTORINNEN UND AUTOREN

- **Esther BAUR** hat Geschichte und Kunstgeschichte studiert und ist seit 2007 Staatsarchivarin des Kantons Basel-Stadt. Ausserdem ist sie Lehrbeauftragte am Departement Geschichte der Universität Basel im Rahmen des Moduls Archive-Medien-Theorie (AMT), Verfasserin diverser Publikationen zu archivwissenschaftlichen Themen und zur historischen Fotografie; Mitglied des Herausgebergremiums der neuen Geschichte des Kantons Basel-Stadt, ferner Vizepräsidentin der Schweizer Archivdirektorenkonferenz und Mitglied der Aufsichtskommission der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST).
- **Flurina CAMENISCH** hat von 2010 bis 2017 Geschichte und Ethnologie an der Universität Zürich studiert. Neben dem Studium arbeitete sie als studentische Mitarbeiterin im Bereich Aktenerschliessung im Staatsarchiv Zürich. Seit 2018 ist sie als wissenschaftliche Mitarbeiterin im Staatsarchiv Graubünden in den Bereichen Überlieferungsbildung und digitale Archivierung tätig.
- **Jürgen ENGE** hat Informatik studiert. Nach Stationen an der Staatlichen Hochschule für Gestaltung Karlsruhe, am Zentrum für Kunst und Medientechnologie Karlsruhe, an der Zürcher Hochschule der Künste, der Hochschule der Künste Bern und der Hochschule für Angewandte Wissenschaften und Kunst Hildesheim/Holzminden/Hannover ist er 2016 als Leiter der ICT/Center for Digital Matter an die Hochschule für Gestaltung und Kunst FHNW Basel gekommen.
- **Paul FLAMME** – Studium der Informatik und Geschichte in Dortmund, Heidelberg und Münster. Seit 1986 als Wissenschaftlicher Mitarbeiter am Staatarchiv Hamburg. Aktuell Leiter des Referats „Erhaltung digitalen Archivguts, Archivische Fachinformationssysteme, Transparenzportal Hamburg“.
- **Christine FRIEDERICH** ist Archivarin beim Landesarchiv Nordrhein-Westfalen in Duisburg. Sie ist dort vor allem für die Behördenberatung zuständig und unterstützt Behörden bei allen Fragen zur digitalen Schriftgutverwaltung und zur E-Akte.
- **Christine GIGLER** – Studium der Geschichte und Germanistik an der Alpen-Adria-Universität Klagenfurt, postgraduale Ausbildung zur Archivarin am Institut für Österreichische Geschichtsforschung der Universität Wien, berufsbegleitendes Masterstudium Archivwissenschaft an der Fachhochschule Potsdam. Nach Tätigkeiten im Kärntner Landesarchiv ab 2003 Archivarin im Wiener Stadt- und Landesarchiv. Seit 2005 Archivarin im Archiv der Erzdiözese Salzburg und dort zuständig für Records Management, Behördenberatung und Planungen im Bereich der digitalen Archivierung.

- **Alexander HERSCHUNG** studierte an der Rheinischen Friedrich-Wilhelms-Universität in Bonn Mathematik und Psychologie (1997). Daran anschließend arbeitete er selbständig im Bereich der Projektentwicklung. Seit 2000 ist er bei startext als Software-Ingenieur und als Leiter der Archiv-Software tätig. 2014 übernahm er das Unternehmen und ist seitdem geschäftsführender Gesellschafter.
- **Pavel IRCING** is the associate professor at the Department of Cybernetics and the New Technologies for the Information Society (NTIS) Research Center of the University of West Bohemia. He holds the Ph. D. in Cybernetics from the same university and has been involved in many projects dealing with automatic speech recognition and information retrieval.
- **Martin KAISER** hat ein Studium in Sprachwissenschaft und Geschichte und ein Nachdiplom Informatik in Basel und Berlin absolviert. Anschliessend an Forschungsprojekten und in der Entwicklung in verschiedenen Informatikfirmen mit dem Schwerpunkt Information Retrieval gearbeitet. Seit 2007 Berater für Digitale Langzeitarchivierung bei der KOST, «Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen».
- **Lambert KANSY** studierte Geschichte, Osteuropäische Geschichte und Soziologie in Basel und Berlin; Fernweiterbildung Archiv an der Fachhochschule Potsdam und Abschluss als Diplom-Archivar (FH). Seit 2000 Wissenschaftlicher Archivar im Staatsarchiv Basel-Stadt; seit 2008 Leiter der Abteilung Informatik und zuständig für den Bereich Archivinformatik.
- **Christian KEITEL** – seit 2000 beim Landesarchiv Baden-Württemberg, derzeit zuständig für Koordination der digitalen Archivierung und der Überlieferungsbildung, seit 2015 Honorarprofessor an der FH-Potsdam.
- **Ulrike KORTE** arbeitet seit 2004 beim Bundesamt für Sicherheit in der Informationstechnik. Sie promovierte im Jahr 1981 im Fach Mathematik an der Universität Münster und arbeitet seitdem in den Bereichen Software-Engineering und Datensicherheit. Ulrike Korte ist (Co-)Author von einem Buch und von mehr als 50 Artikeln für Zeitschriften und Konferenzen und war und ist Mitglied von nationalen und internationalen Standardisierungsgruppen, so z.B. aktuell u.a. von “ISO/TC 154 Long term signature profiles” und European Telecommunications Standards Institute (ETSI).
- **Tomasz KUSBER** arbeitet seit 2017 als wissenschaftliche Mitarbeiter beim Fraunhofer Institut für offene Kommunikationssysteme (FOKUS). Er schloss sein Studium der Informatik mit einem Diplom an der Technischen Universität Berlin ab und arbeitete als wissenschaftliche Mitarbeiter, Software Developer, Leiter einer Entwicklungsgruppe sowie Unternehmensberater im Bereich offene Kommunikationssysteme, Software Engineering, und Datensicherheit. Hr. Kusber ist ein (Co-)Autor der Technischen Richtlinie 03125 (TR-ESOR) des BSI sowie mehreren Publikationen zu diesem Thema.

- **Martin KUNZ** ist seit 2015 Geschäftsführer der scope solutions ag. Zuvor war der studierte Informatiker viele Jahre im Bereich Dokumentenmanagement / Records Management tätig und hat in dieser Funktion auch einen der grössten Schweizer IT Dienstleister in diesem Umfeld geleitet. Sein Themenschwerpunkt ist heute die Beratung im OAIS-Umfeld.
- **Martin LHOTÁK** works from 1997 at the Library of the Czech Academy of Sciences. From 2007 he was for two 5 year terms appointed as the director and currently he is in the position of the Library deputy director responsible for research, development and technologies. At the research level Martin Lhotak focuses on development of open source solutions supporting digitization processes, dissemination and archiving of digital documents – Kramerius, ProArc, ARCLib. He is also engaged in science evaluation, especially in bibliometrics field and he participates in digital humanities projects with concern mainly on development of new tools and building of information infrastructures.
- **Tabea LURK** hat Kunstwissenschaft und Medientheorie studiert. Nach einem Volontariat am Zentrum für Kunst und Medientechnologie Karlsruhe und war sie 2006–2015 im Fachbereich Konservierung und Restaurierung der Hochschule der Künste Bern BFH tätig. Seit 2015 leitet sie die Mediathek der Hochschule für Gestaltung und Kunst FHNW in Basel.
- **Martin LÜTHI** absolvierte eine Ausbildung zum Programmierer, anschließend Weiterbildung zum Wirtschaftsinformatiker in St. Gallen und Zürich. Diverse Tätigkeiten mit Führungsfunktion im Bereich Dokumentenmanagement, Output-Management und digitale Archivierung bei Banken, Informatikdienstleistern und Rechenzentren. Seit 2006 Leiter der Abteilung Aktenführung und digitale Archivierung im Staatsarchiv St. Gallen.
- **Konrad MECKEL** – Diplomstudium der Dokumentation an der Fachhochschule Potsdam, 2010–2012 Stadtarchiv Neu-Ulm, seit 2012 Mitarbeiter des Landesarchivs Thüringen im Projekt Digitales Magazin des Freistaats Thüringen
- **Kai NAUMANN** arbeitet seit 2004 als Archivar beim Landesarchiv Baden-Württemberg, derzeit am Standort Stuttgart. Als Historiker mit einem Hang zur Technik hat er sich intensiv mit der Bewertung, Übernahme, Aufbereitung und Nutzung digitaler Unterlagen beschäftigt. Dem kollegialen Austausch widmet er sich unter anderem im Rahmen des DIMAG-Verbunds, des Kompetenznetzwerks nestor und der KLA (Konferenz der staatlichen Archive in Deutschland).
- **Frank OBERMEIT** – Studium zum Dipl.-Informatiker an der TU Dresden (Abschluss 1993), ab 1976 in Rechenzentren tätig, von 1991 bis 2011 im Landesrechenzentrum Sachsen-Anhalt; seit 2011 als Informatiker im Landesarchiv Sachsen-Anhalt beschäftigt mit Aufgaben der digitalen Archivierung und Leitung des IT-Bereichs.

- **Eliška PAVLÁSKOVÁ** works as an archivist at Charles University and has almost ten years of experience with digital documents and their long-term preservation in academic libraries and archives.
- **Jens PETERS** studierte an der Universität zu Köln Informationsverarbeitung, Politische Wissenschaften und Mittlere und Neuere Geschichte (M. A.). Nach seiner Tätigkeit als Wissenschaftlicher Mitarbeiter am Lehrstuhl Prof. Manfred Thaller arbeitet er seit 2014 für das Systemhaus LVR-InfoKom des Landschaftsverbandes Rheinland (LVR), maßgeblich an der LZA-Lösung „Digitales Archiv NRW Software Suite (DNS)“ im Auftrag des Landes NRW.
- **Svenia POHLKAMP** – Studium der Buchwissenschaft, Deutschen Philologie und Alten Geschichte an der Universität Mainz sowie der Bibliotheks- und Informationswissenschaft an der HU Berlin, seit 2015 Mitarbeiterin des Landesarchivs Thüringen im Projekt Digitales Magazin des Freistaats Thüringen
- **Marta RIESS** hat nach ihrer Promotion in Geschichte an der Universität Wien ein Masterstudium in digitaler Dokumentation („Documentación Digital“) an der Universität Pompeu Fabra in Barcelona absolviert. Ihre Arbeit im archivarischen Umfeld begann sie im Jahre 2006 im Archiv der Familie Pallavicini und war danach über längere Zeit wissenschaftliche Mitarbeiterin im Archiv der Universität Wien. Nach einem kurzen Intermezzo als Provenienzforscherin an der Alberina in Wien, ist sie seit 2010 Archivarin bei der Internationalen Atomic Energy Agency in Wien.
- **Martin SCHLEMMER** ist Archivar beim Landesarchiv Nordrhein-Westfalen in Duisburg. Er ist dort vor allem für die Behördenberatung zuständig und unterstützt Behörden bei allen Fragen zur digitalen Schriftgutverwaltung und zur E-Akte.
- **Steffen SCHWALM** arbeitet als Principal Consultant bei der msg systems ag. Er ist Experte hinsichtlich der Beweissicherheit elektronischer Unterlagen, Vertrauensdienste und digitale Identitäten, Beweiswerterhaltung und Records Management. Er verfügt über mehr als 10 Jahre Erfahrung im Aufbau und Umsetzung entsprechender organisatorischer wie technischer Lösungen in hochregulierten Industrien und beteiligt sich seit langem aktiv in der nationalen und internationalen Standardisierung so z. B. im BSI, ISO und DIN. Steffen Schwalm studierte Informationswissenschaften und publiziert regelmäßig in seinen Fachthemen.
- **Björn STEFFENHAGEN** – Archivstudium (B. A.) & der Informationswissenschaften (M. A.) an der FH Potsdam 2011–2016, seit 2017 Wiss. Archivar im Landesarchiv Sachsen-Anhalt, dort u. a. zuständig für den Aufbau der digitalen Archivierung sowie Magazinpartnerschaften; gleichzeitig Fernstudium im Master Wirtschaftsinformatik an der Universität Duisburg-Essen.
- **Bernhard STÜSSI** studierte in Bern Geschichte und Slavistik und absolvierte den MAS in Archival, Library and Information Science (ALIS). Nach Statio-

nen in den Stadtarchiven Bern und Zofingen sowie dem Staatsarchiv Aargau ist er seit 2016 als wissenschaftlicher Mitarbeiter Überlieferungsbildung und Records Management im Staatsarchiv Graubünden tätig.

- **Karolína ŠIMŮNKOVÁ** – Studium Jura und Geschichte an der Karls Universität, seit 2004 in Nationalarchiv in Prag, dann im Ministerium des Innern, Referat für Archivwesen und Schriftgutverwaltung, seit 2013 in Nationalarchiv an der Stelle der Leiterin der Abteilung der Bestände 1945–1992 und seit 2017 Referentin in der Abteilung 5 – Neueste Bestände ab 1993/ Digitalarchiv.
- **Martin ŠISLER** – Studium Geschichte an der Karls Universität, seit 2007 in Nationalarchiv in Prag als Referent in der Abteilung für Archivalien nach 1992 und elektronische Dokumente.
- **Jan ŠVEC** is the scientific researcher at the New Technologies for the Information Society (NTIS) Research Center of the University of West Bohemia. He is also an employee of the SpeechTech company, where he works on development and deployment of speech-related technologies He holds the Ph. D. in Cybernetics and participates on many projects involving speech technologies.
- **Christine TRÄGER** – Studium der Neueren/Neuesten Geschichte, Kunstgeschichte und Klass. Archäologie an der Universität Jena, 2010–2011 Volontärin beim Bayerischen Rundfunk und postgraduale Ausbildung zur Wiss. Dokumentarin an der Fachhochschule Potsdam mit Schwerpunkt elektronische Archivierung, am Landesarchiv Thüringen seit 2012 Leiterin des Projektes Digitales Magazin des Freistaats Thüringen und seit 2017 Referentin im Referat 6.4 – Neueste Bestände ab 1990.
- **Zdeněk VAŠEK** works as a project manager in the Archives of Charles University where he coordinates the project of creating an Archived Information System for the management and storage of digital documents. Before that he acted as a head of Digital Preservation Standards Department of the National Library of the Czech Republic.
- **Milan VOJÁČEK** absolvierte Archivistik, Geschichte und Philosophie an der Karlsuniversität Prag und gleichzeitig die Tschechische Technische Universität. Seit 2004 arbeitet er in Nationalarchiv Prag als Archivar (zuständig für die Periode 1848–1918), von 2011 Mitglied des Projekts Digitales Archiv, jetzt Abteilungsleiter für Archivalien nach 1992 und elektronische Dokumente.
- **Tobias WILDI** ist Co-Geschäftsführer von docuteam, einem Schweizer Anbieter von Open Source-Software und einer Cloudlösung für digitale Archivierung. Wichtigster Teil seiner Arbeit ist die Entwicklung von Modellen für archivische Metadaten wie „Matterhorn METS“ und neu auch das „Matterhorn RDF Data Model“.

- **Daniel WITTMANN** – Studium der Medieninformatik an der Bauhaus-Universität Weimar, von 2015 bis 2017 Studentische Hilfskraft und seit 2018 Mitarbeiter im Landesarchiv Thüringen im Projekt Digitales Magazin des Freistaats Thüringen.

23. zasedání pracovní skupiny Archivace záznamů z digitálních systémů (AUdS), Praha, 12. a 13. března 2019

Editor: Karolína Šimůnková, Milan Vojáček

Kolektivní monografie vznikla na základě konference Archivierung von Unterlagen aus digitalen Systemen (Archivace záznamů z digitálních systémů) a souvisejících seminářů, které se konaly ve dnech 12. a 13. března 2019 v Národním archivu v Praze.

Vydavatel: Národní archiv, Archivní 2257/4, 149 00, Praha 4

Vydání: první

Tisk: Národní archiv, Archivní 2257/4, 149 00, Praha 4

Odborná redakce: Zbyšek Stodůlka, Karolína Šimůnková, Milan Vojáček

Jazyková redakce: autoři příspěvků

Překlad: Stefan Scholz (čl. Indexace audiovizuálních archivů s využitím metod automatického rozpoznávání řeči a obrazu)

Grafický návrh a sazba: © Eva Tamchynová

© Národní archiv, 2019

www.nacr.cz

ISBN 978- 80-7469-100-3 (pdf)

23. Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen am 12. und 13. März 2019 in Prag

Herausgeber: Karolína Šimůnková, Milan Vojáček, Národní archiv

Postanschrift: Archivní 2257/4, 149 01, Praha 4, Tschechische Republik

Herausgabe: 1.

Druck: Národní archiv, Archivní 2257/4, 149 01, Praha 4

Redaktion: Zbyšek Stodůlka, Karolína Šimůnková, Milan Vojáček

Korrektorat: Autorinnen und Autoren

Übersetzung: Stefan Scholz (Artikel: Die Indexierung von audiovisuellen Archiven unter Benutzung von Methoden der automatischen Sprach- und Bilderkennung)

Grafik, Satz: © Eva Tamchynová

© Národní archiv, 2019

(Alle Rechte für Beiträge liegen bei den jeweiligen Autorinnen und Autoren)

www.nacr.cz

ISBN 978- 80-7469-100-3 (pdf)

