

Digitale Bestandserhaltung in DIMAG

Transferarbeit

im Rahmen

der Laufbahnprüfung für den höheren Archivdienst
an der Archivschule Marburg

Laura Nippel

Landesarchiv Baden-Württemberg

53. Wissenschaftlicher Lehrgang

Vorgelegt am 27. März 2020

Gutachter:

Prof. Dr. Christian Keitel (Archivschule Marburg)

Dr. Kai Naumann (Landesarchiv Baden-Württemberg)

Inhalt

| | | |
|-------|--|----|
| I. | Einleitung | 1 |
| II. | Doppelte Beobachtung: Community Watch und Technology Watch | 3 |
| III. | Das Hilfsmittel Formatregister | 5 |
| IV. | Formate für die digitale Langzeitarchivierung | 7 |
| V. | Formaterkennung und -validierung | 12 |
| VI. | Formate in DIMAG – eine kleine Bestandsaufnahme | 19 |
| VII. | Das Datenmodell der Preservation Action Registries (PAR)..... | 22 |
| VIII. | Protokollierung von Erhaltungsmaßnahmen in DIMAG | 26 |
| IX. | Eine Ampel für bedrohte Formate..... | 29 |
| X. | Schlussbetrachtungen | 32 |
| | Zusammenfassung | 34 |
| | Anhang | 35 |
| | Anlage 1: Archivierungsfähige Dateiformate | 35 |
| | Anlage 2: PAR Conceptual Model..... | 39 |
| | Anlage 3: PAR Business Rule Eintrag | 40 |
| | Anlage 4: PAR Preservation Action Eintrag..... | 41 |
| | Abkürzungsverzeichnis | 42 |
| | Literatur- und Quellenverzeichnis..... | 43 |
| | Gesprächsvermerke | 49 |

I. Einleitung

Zur Vermeidung eines „Digital Dark Age“¹ bedarf es einer aufmerksamen Beobachtung der Umwelt. Deren Wandel kann die Auswertbarkeit der Informationen sowohl positiv als auch negativ beeinflussen. Technische Innovationen bieten auf der einen Seite neue Möglichkeiten zur Erschließung, Erhaltung und Nutzung digitaler Informationen, auf der anderen Seite können bisher bekannte und genutzte Verfahren, Programme und Formate unbrauchbar werden. Insbesondere die Obsoleszenz von Formaten steht im Zentrum der digitalen Bestandserhaltung.² Dem „Aussterben“ von Dateiformaten wird die Sicherung der Les-, Interpretier- und Darstellbarkeit der Informationen entgegengestellt. Die Aufgabe der digitalen Langzeitarchivierung ist nicht der Erhalt der ursprünglichen Dateien und Formate, sondern die Sicherung und Zugänglichmachung digital gespeicherter Informationen unabhängig von technischen Veränderungen und orientiert an den antizipierten Wünschen und Kenntnissen künftiger Nutzer, der Designated Community.³ Transparenz und Nachvollziehbarkeit der Entscheidungen sind wichtige Faktoren für das Vertrauen der Nutzer in die Authentizität und Glaubwürdigkeit der Informationen.⁴

Nachdem der Aufbau digitaler Archivinfrastrukturen mittlerweile im Kern abgeschlossen ist, rückten in den vergangenen Jahren die digitalen Erhaltungsmaßnahmen verstärkt in den Blick. Die Bestandserhaltungsentscheidungen im Digitalen unterscheiden sich durch ihre kontinuierliche Repetition von jenen im Analogen: „An die Stelle der passiven Bestandserhaltung tritt eine aktive Bestandserhaltung.“⁵ Diese

¹ Vgl. Griffin, Andrew, ‚Digital Dark Age‘ Could Leave Historians With No Records of the 21st Century, in: Independent 13.02.2015, <https://www.independent.co.uk/life-style/gadgets-and-tech/news/digital-dark-age-could-leave-historians-with-no-records-of-the-21st-century-10043516.html> (19.03.2020).

² Zur Obsoleszenzproblematik vgl. u.a. DPC, Digital Preservation Handbook, 2. Aufl. 2015, <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards> (03.03.2020).

³ Vgl. nestor-Arbeitsgruppe Digitale Bestandserhaltung, Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 2.0 (2012), <http://nbn-resolving.de/urn:nbn:de:0008-2012092400> (22.01.2020), S. 3; Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS). Magenta Book (CCSDS 650.0-M-2) (2012), <https://public.ccsds.org/Pubs/650x0m2.pdf> (15.01.2020), S. 1-1. Das Konzept der Designated Community entspringt dem OAIS: „An identified group of potential Consumers who should be able to understand a particular set of information.“, ebd., S. 1-11.

⁴ Vgl. Keitel, Christian, Zwölf Wege ins Archiv. Umriss einer offenen und praktischen Archivwissenschaft, Stuttgart 2018, S. 140f.; Ders., Authentische Archive. Wunsch und Wirklichkeit, in: Herrmann, Tobias (Hrsg.), Verlässlich, richtig, echt – Demokratie braucht Archive! 88. Deutscher Archivtag in Rostock, Fulda 2019, S. 123–131.

⁵ Ders., Der nestor-Leitfaden zur Digitalen Bestandserhaltung und seine Folgen für die Archive, in: Ders. (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von

beginnt schon vor der Übernahme und endet nicht mit der ersten Maßnahme, sondern erfordert ein beständiges Eingreifen und grundlegende Konzepte, die unkoordinierte vereinzelte Erhaltungsmaßnahmen verhindern und planvolle, nachvollziehbare und bestenfalls automatisierte Verfahren festlegen.⁶

Während die analoge Bestandserhaltung insbesondere auf Veränderungen in ihren Archivalien, an den ‚Datenträgern‘ selbst reagiert, reagiert die digitale auf Veränderungen der Außenwelt, auf den technischen Fortschritt sowie das Verhalten der Nutzer. Es bedarf somit einer systematischen Beobachtung der Umwelt, die sich in Anbetracht des unübersichtlichen Feldes herausfordernd gestaltet. Diese Arbeit kann weder ein vollumfängliches Konzept der digitalen Bestandserhaltung noch Antworten auf alle offenen Fragen liefern. Stattdessen sollen grundlegende Problematiken der Bestandserhaltung von Dateiformaten angesprochen, bestehende Hilfsmittel vorgestellt, neue Konzepte diskutiert und Vorschläge für praktische Umsetzungen im DIMAG-Umfeld gemacht werden, um einen Überblick über die Dynamiken zu erlangen. Erkenntnisleitend ist dabei die Frage, welche Maßnahmen in DIMAG schon umgesetzt werden und welche weiteren notwendig sind, um die Obsoleszenz von Dateiformaten und damit den Informationsverlust in den aktuellen Repräsentationen zu verhindern. Dabei kommt insbesondere der Vernetzung innerhalb der Bestandserhaltungscommunity eine große Bedeutung zu. Nachdem zunächst die Beobachtungsaufgaben umrissen werden, folgt ein Überblick über bestehende Formatregister. An die Auseinandersetzung mit langzeitarchivierungsfähigen Dateiformaten schließt ein Abriss der Verfahren und Tools zur Formaterkennung und -validierung an. Aus einer kleinen Auswertung der DIMAG-Gesamtstatistik sollen Rückschlüsse auf bisherige Abfragemöglichkeiten in DIMAG gezogen werden. Darüber hinaus wird die Anschlussfähigkeit des Projektes zu Preservation Action Registries (PAR) diskutiert und Vorschläge für eine Protokollierung von Bestandserhaltungsmaßnahmen in DIMAG unterbreitet. Abschließend soll die Idee eines Ampelsystems zur Kennzeichnung der Obsoleszenzgefahr von Formaten skizziert werden. Die Arbeit konzentriert sich vornehmlich auf die technischen Veränderungen in der Umwelt eines digitalen Archivs.

Unterlagen aus Digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“, Stuttgart 2013, S. 267–280, hier S. 269.

⁶ Vgl. ebd.

II. Doppelte Beobachtung: Community Watch und Technology Watch

Digitale Bestandserhaltung zeichnet sich durch eine doppelte Beobachtungsaufgabe aus: Nicht nur technische Veränderungen werden überwacht, sondern Änderungen im Verhalten und in den Anforderungen der Nutzer. In der ‚heiligen Schrift‘ der digitalen Langzeitarchivierung, dem Reference Model for an Open Archival Information System (OAIS)⁷ sind sechs grundlegende Funktionseinheiten definiert, zu denen als übergreifende Einheit die Erhaltungsplanung (Preservation Planning) gehört. Bestandteil des Preservation Planning ist neben der Beobachtung der Umwelt des OAIS auch die Entwicklung von Strategien und Plänen, welche die Lesbarkeit und das Verständnis der gespeicherten Informationen durch die Designated Community sicherstellen sollen.⁸ Die Beobachtung wird in die Bereiche „monitor technology“ und „monitor designated community“ aufgespalten.

Die Funktion „monitor designated community“ beinhaltet die konstante Überwachung von Veränderungen innerhalb der vorgesehenen Zielgruppe. Dies fordert eine gesteigerte Aufmerksamkeit für sich wandelnde Erwartungen der Nutzer an die Dienstleistungen des Archivs, für ihr Nutzungsverhalten und für ihre technischen Kompetenzen. Archive müssen sich somit ständig fragen, welche Dateiformate und Softwareprodukte von der Designated Community tatsächlich genutzt werden. Die Bewertung kann dabei je nach Archivaliengattung und Nutzergruppe unterschiedlich ausfallen. Die angemessene und differenzierte Einschätzung des (technischen) Wissens der Nutzer kann nur in einem ständigen Kommunikationsprozess mit der Designated Community vorgenommen werden. Denkbar sind Workshops, Umfragen und Einzelgespräche mit Nutzern.⁹

Die Funktion „monitor technology“ dient der Beobachtung technischer Entwicklungen und Neuerungen im Bereich digitaler Technologien, Informationsstandards sowie Hard- und Software, um rechtzeitig auf Veränderungen reagieren zu können, welche die Nutzbarkeit von Informationen gefährden könnten. Die Obsoleszenz

⁷ Das OAIS-Referenzmodell wurde 2002 vom Consultative Committee of Space Data Systems (CCSDS) unter Führung der NASA in einer ersten, 2012 in einer zweiten Version veröffentlicht. Im Folgenden folgen die Belege der englischen Ausgabe, während sich die deutschen Begrifflichkeiten an der von nestor herausgegebenen deutschen Übersetzung orientieren: nestor, Referenzmodell für ein Offenes Archiv-Informationssystem. Deutsche Übersetzung 2.0 (2013), <https://nbn-resolving.org/urn/resolver.pl?urn=urn:nbn:de:0008-2013082706> (15.01.2020). Zur Entstehungsgeschichte von OAIS vgl. Keitel, Zwölf Wege, S. 48–53. Zur Kritik am OAIS vgl. Harvey, Ross/Weatherburn, Jaye, Preserving Digital Materials, 3. Aufl., Lanham/London 2018, S. 78f.

⁸ Vgl. Consultative Committee for Space Data Systems, OAIS (Magenta Book), S. 4-1f.

⁹ Vgl. ebd., S. 4-14.

von Dateiformaten ist dabei eines der größten Risiken, weitere Gefahren liegen aber auch im Alterungsprozess von Speichertechniken oder Hard- und Software.¹⁰

Der 2013 erschienene Leitfaden zur digitalen Bestandserhaltung des deutschen Netzwerks nestor stärkt die vom OAIS postulierten Beobachtungsaufgaben, jedoch unter den eingängigeren Titeln „Community Watch“ und „Technology Watch“. Der Leitfaden legt einen Schwerpunkt auf die Community Watch, da er auch Technologieeinschätzungen konsequent mit Blick auf den „wesentlichen Referenzpunkt“¹¹, die Designated Community, vornimmt. Die Entscheidung für oder gegen eine Erhaltungsmaßnahme sollte immer an den Anforderungen und Möglichkeiten der gedachten Zielgruppe gemessen werden.¹² Die Technology Watch scheint dagegen etwas in den Hintergrund zu treten. Eng an das OAIS angelehnt formuliert auch der Leitfaden die Aufgabe, Standards und technologische Entwicklungen aufmerksam zu beobachten.¹³ Sowohl OAIS als auch der nestor-Leitfaden beschreiben die Beobachtungsaufgaben auf einer sehr abstrakten Ebene, ohne dabei eine Anleitung für eine systematische Durchführung zu bieten.¹⁴

Der anerkannte Metadatenstandard für die Langzeitarchivierung PREMIS¹⁵ hält weder in seinem Datenmodell noch in seinem Data Dictionary explizite Beobachtungsaufgaben fest. Das Datenmodell beschreibt lediglich, welche Informationen zur Umgebung eines Objektes („environment“), beispielsweise Soft- und Hardware zur Darstellung, aufgezeichnet werden sollen.¹⁶ Das Data Dictionary sieht zwar die semantische Einheit des „preservation level“ vor, das Aussagen über den Umfang und

¹⁰ Vgl. ebd., S. 4-14–4-16. Die aus den Monitoring-Funktionen gewonnenen Erkenntnisse haben auch Auswirkungen auf die Verwaltung der Metadaten im OAIS. Technische Neuerungen, aber auch Veränderungen in dem Grundwissen der Designated Community können Anpassungen der Representation Information (Repräsentationsinformation) und der Preservation Description Information (Erhaltungsmetadaten) nach sich ziehen, um das Verständnis der Informationen weiterhin zu gewährleisten. Vgl. ebd., S. 3-4f.

¹¹ nestor-Arbeitsgruppe Digitale Bestandserhaltung, Leitfaden, S. 29.

¹² Vgl. ebd., S. 24f.

¹³ Vgl. ebd., S. 30.

¹⁴ Vgl. Becker, Christoph/Duretec, Kresimir/Petrov, Petar/Faria, Luis/Ferreira, Miguel/Ramalho, Jose Carlos, Preservation Watch. What to Monitor and How, in: Proceedings of iPRES 2012 9th International Conference on Preservation of Digital Objects (2012), http://publik.tuwien.ac.at/files/PubDat_213865.pdf (14.02.2020).

¹⁵ Preservation Metadata: Implementation Strategies; PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata. Version 3.0 (2015), <http://www.loc.gov/standards/premis/v3/index.html> (20.01.2020).

¹⁶ Vgl. ebd., S. 14f.; Caplan, Priscilla, PREMIS verstehen (2017), https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/berichte/pPREMISverstehen2017.pdf?__blob=publicationFile&v=1 (20.01.2020), S. 10f.

die Art anzustrebender Erhaltungsmaßnahmen tätig, jedoch nicht den Auslöser für diese festhält und bewertet.¹⁷

III. Das Hilfsmittel Formatregister

So wichtig die Beobachtungsaufgaben sind, so schwer ist es, sie gezielt und kontinuierlich umzusetzen. Nicht nur dem Landesarchiv Baden-Württemberg fehlt es bisher an einem Überwachungssystem. Existieren im Bereich der Technology Watch immerhin einige, wenn oft auch lückenhafte Hilfsmittel, so scheint die Beobachtung der Nutzer derzeit allein auf der Einschätzung allgemeiner gesellschaftlicher Trends und Eindrücke zu fußen, die am Rande der Nutzer- oder Behördenberatung, wenn nicht auch aus dem privaten Umfeld gewonnen wurden.¹⁸

Für eine Technology Watch im Bereich der Formate kann auf Formatregister als Hilfsmittel zurückgegriffen werden.¹⁹ PRONOM stellt die bekannteste und am stärksten genutzte Formatregistry dar. Sie wurde 2002 vom britischen Nationalarchiv zunächst als internes Hilfsmittel, seit 2004 als öffentlich zugängliche Webressource entwickelt und umfasst mittlerweile mehr als 1.300 Dateiformate.²⁰ Damit handelt es sich um eine sehr umfangreiche Registry, die dennoch von einer Vollständigkeit der verzeichneten Formate als auch der Informationen weit entfernt ist. Die National Archives pflegen mit Unterstützung aus der Community die Einträge, die unter anderem technische Informationen über die Struktur, über den Entwickler und verwandte Formate enthalten. Auch werden der MIME-Type und die Signature bzw. Magic Number,

¹⁷ Vgl. PREMIS Editorial Committee, *Data Dictionary*, S. 42–49; Caplan, *PREMIS verstehen*, S. 19f.

¹⁸ Man steht hier vor denselben Problemen wie bei der Bewertung oder bei der Bestimmung signifikanter Eigenschaften. Das Schweizerische Bundesarchiv verfolgt bspw. mit dem Konzept der „partizipativen Bewertung“ das Ziel, die Nutzer stärker in die Bewertungsentscheidungen einzubinden, vgl. Schweizerisches Bundesarchiv BAR, *Partizipative Bewertung im BAR. Auswertung Pilot und Überführung Betrieb. Version 1.0* (2019), <https://www.bar.admin.ch/dam/bar/de/dokumente/diverses/Bericht%20BAR%20Pilot%20partizipative%20Bewertung.pdf.download.pdf/Bericht%20Auswertung%20Pilot%20partizipative%20Bewertung%20BAR.pdf> (21.02.2020). Ein überraschender Erfolg war zuletzt eine kleine Twitter-Umfrage von Mareike König (DHI Paris) über die allgemeinen Wünsche der Nutzer ans Archiv, vgl. König, Mareike, *Was sich Historiker*innen von Archiven wünschen: eine Umfrage* (2019), <https://dhdhi.hypotheses.org/6107> (05.11.2019). Antwort aus der Archivwelt: Worm, Peter, *Nur ein Wunschkonzert? Hinweis auf eine Twitterumfrage und ihre Auswertung* (2019), <https://archivamt.hypotheses.org/11487> (05.11.2019). S.a. Keitel, *Zwölf Wege*, S. 223–239; Ders., *Benutzerinteressen annehmen und signifikante Eigenschaften festlegen. Einige neue Aufgaben für Archivare*, in: Schmitt, Heiner (Hrsg.), *Archive im digitalen Zeitalter. Überlieferung, Erschließung, Präsentation*. 79. Deutscher Archivtag in Regensburg, Neustadt a.d. Aisch 2010, S. 29–42.

¹⁹ Nicht nur bei den Formaten, sondern auch bei den Datenbanken handelt es sich um ein undurchsichtiges Feld, weshalb hier kein Anspruch auf Vollständigkeit erhoben wird.

²⁰ Vgl. <https://www.nationalarchives.gov.uk/aboutapps/PRONOM/default.htm> (23.02.2020). Derzeit wird daran gearbeitet, die PRONOM-Datenbank in eine linked data Version zu überführen. Die Planungen und internen Diskussionen sind derzeit noch nicht abgeschlossen, weshalb noch kein Zeitplan für die Umsetzung veröffentlicht wurde, vgl. Gespräch mit David Clipsham (PRONOM, The National Archives), Nicola Wissbrock (The National Archives), Kai Naumann (LA BW), 17.02.2020.

also das charakteristische Bitmuster, festgehalten. Durch den PRONOM Persistent Unique Identifier (PUID) steht für jeden Formateintrag ein eindeutiger Identifier zur Verfügung, wodurch die Unzulänglichkeiten bisheriger Erkennungsmerkmale wie der Dateinamenserweiterung oder dem MIME-Type durch die Eindeutigkeit und angemessene Granularität der PUID ausgeglichen werden sollen.²¹ Die PUID hat weitgehend Anerkennung gefunden; sie wird sowohl von einigen Identifizierungswerkzeugen verwendet, als auch in die Metadaten in DIMAG aufgenommen.

Die Library of Congress (LOC) stellt auf ihrer Website „The Sustainability of Digital Formats“²² ein Register mit sehr ausführlichen Beschreibungen der Formate, Einschätzungen zu deren Archivfähigkeit und Handlungsempfehlungen zur Verfügung. Es erreicht zwar nicht den Umfang von PRONOM, die Zahl von rund 500 Einträgen ist dennoch beachtlich und die Informationsdichte oftmals höher als jene in PRONOM. Für eine umfassende Informierung über Formate und zur Planung von Bestandserhaltungsschritten sollte neben PRONOM die Datenbank der LOC hinzugezogen werden.

In den letzten zwanzig Jahren gab es weitere Initiativen zum Aufbau von Formatdatenbanken, die sich jedoch aufgrund von Projektstrukturen und -finanzierung als kurzlebig herausstellten. Lediglich drei von neun im nestor-Handbuch 2010 beispielhaft aufgeführten Registries sind noch im Netz zu finden.²³ Sowohl die Global Digital Format Registry (GDFR) der Harvard University Libraries als auch der Versuch, die GDFR mit PRONOM in der Unified Digital Format Registry (UDFR) der California Digital Library zu verbinden, sind gescheitert.²⁴ Des Weiteren besteht das Wiki eines Crowdsourcing Registers „Let’s Solve the File Format Problem!“, das vom

²¹ Vgl. <https://www.nationalarchives.gov.uk/aboutapps/pronom/puid.htm> (24.02.2020). Zum Aufbau der PUIDs s.a. <https://wiki.dnb.de/display/NESTOR/Pronom%3A+Persistenz+von+PUIDs> (24.02.2020).

²² https://www.loc.gov/preservation/digital/formats/fdd/browse_list.shtml (24.02.2020). Zum Aufbau der Datenbank vgl. Arms, Caroline/Fleischhauer, Carl, Digital Formats. Factors for Sustainability, Functionality and Quality (2005), https://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf (24.02.2020).

²³ Es funktionieren nur die Links zu PRONOM, zur Library of Congress und zu FILEExt, einer Datenbank zu Dateinamenserweiterungen, vgl. Aschenbrenner, Andreas/Wollschläger, Thomas, File Format Registries, in: Neuroth, Heike/Oßwald, Achim/Scheffel, Regine/Huth, Karsten (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010), <http://nbn-resolving.de/urn:nbn:de:0008-2010071949> (28.01.2020), Kap. 7:19–Kap. 7:22, hier Kap. 7:20.

²⁴ Die Arbeit am GDFR lief 2003–2008, ohne dass am Ende eine Datenbank veröffentlicht wurde. Eine Website ist nicht mehr zu finden. UDFR wurde von 2010–2012 durch die Library of Congress mit einer Projektförderung versehen, die Registry 2016 aber eingestellt. Auf der verbliebenen Website wird nun auf PRONOM verwiesen, vgl. <http://www.udfr.org/> (24.02.2020). S.a. Corrado, Edward M./Moulaison Sandy, Heather, Digital Preservation for Libraries, Archives, and Museums, 2. Aufl., Lanham u.a. 2017, S. 207f.; Brown, Adrian, Practical Digital Preservation. A How-To Guide for Organizations of Any Size, London 2013, S. 279.

sogenannten Archive Team betrieben wird. Inwieweit die Datensätze des Wikis noch gepflegt werden, ist unklar. Neben der Dateinamenserweiterung, dem MIME-Type und der PUID von PRONOM wird für die technischen Informationen hauptsächlich mit Verlinkungen auf Software, Programmcodes oder Formatspezifikationen gearbeitet.²⁵

In Anbetracht der Kurzlebigkeit vieler Initiativen ist die Beständigkeit von PRONOM nicht hoch genug einzuschätzen. Dies ist insbesondere mit der PUID und deren maschineller Verarbeitung zu erklären, was keines der anderen Formatregister bieten konnte. Ein automatisiertes Warnsystem stellt jedoch auch PRONOM nicht dar. Die Ausführung der Technology Watch in Bezug auf Dateiformate ist bisher weiterhin davon abhängig, wie stark ein Archiv oder der einzelne Archivar in den Diskurs und in die Community der digitalen Bestandserhaltung eingebunden ist. Netzwerke, Organisationen und Ausschüsse wie beispielsweise nestor, die Open Preservation Foundation (OPF)²⁶ oder der KLA Ausschuss „Digitale Archive“ versuchen die nötige Austauschplattform dafür zu bieten. Die Bestandsaufnahme zu den Formatregistern hat gezeigt, dass nicht nur die Dynamiken auf dem Feld der Formate, sondern auch auf dem der Registries beobachtet werden muss. Dabei sollte auf Kooperation mit anderen Institutionen gesetzt werden. Langfristig erfolgversprechend ist nur die Verankerung bei einer etablierten Institution, weshalb die Beteiligung der nestor-AG „Formaterkennung“ an der Pflege und Weiterentwicklung der PRONOM-Datenbank der richtige Weg ist.²⁷

IV. Formate für die digitale Langzeitarchivierung

Über die Anforderungen und Auswahlkriterien von Archivierungsformaten wurde in den vergangenen Jahren eine ausführliche internationale Debatte geführt, die hier nicht

²⁵ Vgl. <http://fileformats.archiveteam.org/> (24.02.2020). Das Wiki ist 2012 aus einem Projekt entstanden, in dem Jason Scott die Öffentlichkeit aufgerufen hatte, innerhalb eines Monats Informationen zu Formaten und Software in einem Wiki zusammenzutragen. Für den Aufruf zum „Let’s Just Solve the Problem Month“ (November 2012) vgl. Scott, Jason, Let’s Just Solve the Problem Month (02.07.2012), <http://ascii.textfiles.com/archives/3645> (24.02.2020). S.a. Corrado/Moulaison Sandy, Digital preservation, S. 208f.

²⁶ Die OPF (2010 als Open Planets Foundation gegründet) ist aus einem EU-Projekt zur digitalen Bestandserhaltung, dem PLANETS-Projekt hervorgegangen, das 2010 endete. Die OPF, zu deren internationalen Mitgliedern Nationalbibliotheken, Archive, Universitäten und Forschungseinrichtungen gehören, engagiert sich auf dem Feld der digitalen Bestandserhaltung (sowohl auf dem Feld der Kompetenzvermittlung als auch der Softwareentwicklung). Vgl. <https://openpreservation.org/about/organization/> (25.02.2020); Harvey/Weatherburn, Preserving Digital Materials, S. 178f.

²⁷ Die AG „Formaterkennung“ arbeitet PRONOM zu und hat eine deutschsprachige Anleitung zur Erstellung von Signatures für PRONOM erstellt: <https://wiki.dnb.de/pages/viewpage.action?pageId=115213928> (24.02.2020).

umfassend nachgezeichnet werden soll, sondern deren Ergebnisse lediglich zusammengefasst werden.²⁸ Es herrscht weitgehender Konsens über die Kriterien zur Bestimmung der Archivierungsfähigkeit eines Formats. Ein anerkannter und im deutschsprachigen Raum weitverbreiteter Anforderungskatalog wurde von der schweizerischen Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) erstellt.²⁹ Dieser stimmt in weiten Teilen mit anderen Katalogen wie beispielsweise dem der LOC oder des Arbeitskreises „Elektronische Archivierung“ der Vereinigung deutscher Wirtschaftsarchive (AKEA) überein.³⁰

Formate sollten über eine öffentliche Spezifikation verfügen, damit ihre Struktur nachvollzogen, geeignete Wiedergabeprogramme und Validierungstools ausgewählt oder die Daten migriert werden können. Darüber hinaus sind Open-Source-Formate proprietären vorzuziehen und wenn möglich solche zu wählen, für die auch unabhängige Open-Source-Implementierungen zur Verfügung stehen. Die Abhängigkeit von bestimmter Hard- oder Software sollte vermieden werden. Eine weite Verbreitung wiederum verspricht, dass es ein größeres Interesse von Entwicklerseite oder anderen engagierten Programmierern gibt, das Format lange nutzbar zu halten, weshalb die Wahrscheinlichkeit geeigneter Konvertierungstools steigt und von längeren Migrationszeiträumen auszugehen ist. Von großer Bedeutung ist der Erhalt aller signifikanten Eigenschaften durch das neue Format und somit die Sicherung der Funktionalität.³¹ Bei aller Funktionalität sollten die Formate aus betriebswirtschaftlichen Gründen über eine hohe Speicherdichte verfügen und somit möglichst wenig Speicherplatz verbrauchen. Weiter sollten sie zur Selbstdokumentation von deskriptiven, technischen oder administrativen Metadaten in der Lage sein.³² Grundsätzlich sind unkomprimierte Formate komprimierten vorzuziehen und es sollten keine Verschlüsselungen oder

²⁸ Eine Zusammenfassung des englischsprachigen Diskurses Anfang der 2000er-Jahre bietet Todd, Malcom, Technology Watch Report. File formats for preservation (2009), http://www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation (07.02.2020), S. 13–17.

²⁹ Vgl. <https://kost-ceco.ch/cms/kriterienkatalog.html> (26.02.2020).

³⁰ Vgl. <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml> (26.02.2020); Arms/Fleischhauer, Digital Formats; Gutzmann, Ulrike/Kamp, Ulrich/Keitel, Christian/Scheidung, Antje, Praktische Lösungsansätze zur Archivierung digitaler Unterlagen. „Langzeitarchivierung“ und dauerhafte Sicherung der digitalen Überlieferung, in: Archiv und Wirtschaft. Zeitschrift für das Archivwesen der Wirtschaft 40 (2007), S. 20–27. Die KOST wiederum stellt eine ausführliche Bibliographie zur Verfügung: <https://kost-ceco.ch/cms/bibliografie.html> (26.02.2020). Für eine tabellarische Übersicht weiterer Anforderungskataloge, die weitgehend mit den hier ausgewerteten übereinstimmen vgl. Todd, Technology Watch Report, S. 38.

³¹ Zur Priorisierung von Dateiformaten anhand der Erhaltung signifikanter Eigenschaften vgl. Keitel, nestor-Leitfaden, S. 276.

³² Die KOST wendet dieses Kriterium derzeit nur auf Videoformate an, während es LOC und AKEA auf alle Formate beziehen.

ähnliche technische Schutzmechanismen die Lesbarkeit einschränken oder verhindern. Die KOST erweitert dieses bewährte Kriterienfeld um bisherige Erfahrungswerte im Archiv mit einem Format (Best-Practice-Analyse), einer Einschätzung der Zukunftsfähigkeit und eine Klassifizierung anhand des Lebenszyklus eines Formats (von altbekannt bis zukünftig), welche die Interpretation der Bewertungsergebnisse unterstützen soll.³³

Die Kataloge stellen Entscheidungshilfen dar, wobei die Gewichtung der unterschiedlichen Kriterien je nach Struktur der Bestände oder Ausrichtung des Archivierungskonzeptes unterschiedlich ausfallen kann. Erst im Anschluss an die Reflexion der Auswahlkriterien sollte eine Festlegung konkreter Archivierungsformate erfolgen. Dabei sollte man sich vergegenwärtigen, dass es sich nicht um eine abschließende Auswahl handelt, sondern um ein dynamisches Verfahren, in dem der Kanon der Archivierungsformate beständig überprüft und angepasst werden muss. Es handelt sich somit um eine Technology Watch, die die Designated Community insbesondere in Bezug auf Verbreitung und Funktionalität des Formats nicht aus den Augen verlieren darf.

Idealerweise sollte jedes Archiv im Rahmen seines Langzeitarchivierungskonzeptes festlegen, welche Formate für die Langzeitarchivierung unterschiedlicher Informationstypen verwendet werden sollten und diese Whitelist in regelmäßigen Abständen aktualisieren. Eine Durchsicht der zugänglichen Informationen auf den Websites des Bundesarchivs und der Archivverwaltungen der Länder hat ergeben, dass nur Hessen, Nordrhein-Westfalen, Rheinland-Pfalz und Thüringen spartenübergreifende Empfehlungen für archivierungstaugliche Formate öffentlich zur Verfügung stellen.³⁴ Es ist davon auszugehen, dass andere Archivverwaltungen solche Listen für den internen Gebrauch pflegen, vermutlich aber nicht alle.

Zur Bewältigung dieser aufwendigen Beobachtungsaufgabe stehen verschiedene Empfehlungslisten bereit, die von nationalen wie internationalen Archiven und Bibliotheken publiziert wurden. Aus dem archivischen Umfeld hat sich erneut die KOST hervorgetan, die in ihrem Katalog archivischer Dateiformate (KaD)

³³ Vgl. https://kost-ceco.ch/cms/kad_intro_de.html (26.02.2020).

³⁴ Kai Naumann hat für das LA BW einen ersten Vorschlag für eine Whitelist unterbreitet: Naumann, Kai, Zwölf Jahre Lernen aus der Praxis. Überlieferungsbildung aus genuin digitalen Unterlagen beim Landesarchiv Baden-Württemberg, in: *Scrinium* 69 (2015), S. 115–136, hier S. 132–134. Ansonsten stellt das LA BW nur Vorgaben für Archivierungsformate von Datentabellen und für PDF/A als Archivierungsformat auf der Website zur Verfügung. Vgl. <https://www.landearchiv-bw.de/web/46913> (02.03.2020).

Empfehlungen zur Verwendung von Dateiformaten für die Langzeitarchivierung anhand ihrer aufgestellten Anforderungskriterien ausspricht.³⁵ Ohne den Anspruch auf Vollständigkeit wurden die Ergebnisse der KOST mit weiteren Empfehlungen verglichen (vgl. Anlage 1). Dabei ist zu beachten, dass diese nicht immer explizit als Langzeitarchivierungsformate gekennzeichnet sind, sondern oftmals als Übernahmeformate. Da die Einrichtungen aufgrund des Aufwands jedoch vermeiden wollen, Formate zu übernehmen, die nicht langzeitarchivierungstauglich sind, sondern direkt nach der Übernahme oder kurze Zeit später migriert werden müssten, ist davon auszugehen, dass nur solche Formate für die Übernahme vorgegeben werden, die sich als Archivierungsformate eignen.³⁶ Neben dem KaD wurden die Kataloge folgender Einrichtungen ausgewertet: Schweizerisches Bundesarchiv³⁷, ETH Zürich³⁸, LOC³⁹, Harvard University Library⁴⁰, Library and Archives Canada⁴¹, National Archives (USA)⁴², National Archives (UK)⁴³, Landesarchiv Nordrhein-Westfalen⁴⁴, Hessisches Landesarchiv⁴⁵, Landesarchiv Rheinland-Pfalz⁴⁶, Landesarchiv Thüringen⁴⁷. Die Auswahl der

³⁵ Die KOST hat bisher 52 Formate einer ausführlichen und differenzierten Bewertung unterzogen. Eine Zusammenfassung der Empfehlungen mit Verweisen auf die detaillierten Analysen bietet https://kost-ceco.ch/cms/kad_recommendation_de.html (28.02.2020). Für eine Übersicht in Form einer Bewertungsmatrix vgl. <https://kost-ceco.ch/cms/Bewertung.html> (28.02.2020).

³⁶ Vgl. Schweizerisches Bundesarchiv BAR, Standards für die Archivierung digitaler Unterlagen. Archivtaugliche Dateiformate (2018), https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf (27.02.2020), S. 3.

³⁷ Vgl. ebd.

³⁸ Die Empfehlungen beziehen sich auf die Ablieferung von Unterlagen für das Repositorium der Research Collection, <https://documentation.library.ethz.ch/display/RC/Archivtaugliche+Dateiformate#space-menu-link-content> (28.02.2020).

³⁹ Vgl. <https://www.loc.gov/preservation/resources/rfs/TOC.html> (28.02.2020).

⁴⁰ Vgl. <https://wiki.harvard.edu/confluence/display/digitalpreservation/Formats+Supported+by+the+DRS> (28.02.2020).

⁴¹ Vgl. <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx> (28.02.2020).

⁴² Vgl. <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html> (28.02.2020).

⁴³ Vgl. <https://www.nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer/> (28.02.2020).

⁴⁴ Die Empfehlungen beziehen sich auf Formate im Rahmen der elektronischen Aktenführung, Landesarchiv Nordrhein-Westfalen, Leitfaden für langzeitstabile Datenformate in der elektronischen Aktenführung. Version 1.1 (2015), https://www.archive.nrw.de/lav/abteilungen/fachbereich_grundsaeetze/BilderKartenLogosDateien/Behoerdeninformation/Bausteine_III_Langzeitstabile_Datenformate.pdf (01.03.2020).

⁴⁵ Vgl. Hessisches Landesarchiv, Archivfähige Dateiformate und ihre Spezifikationen (2016), https://landesarchiv.hessen.de/sites/landesarchiv.hessen.de/files/content-downloads/Archivf%C3%A4hige_Dateiformate_%28Stand_August_2016%29.pdf (02.03.2020).

⁴⁶ Vgl. https://www.landeshauptarchiv.de/fileadmin/user_upload/Gemeinsame_Dateien/Liste_der_archivf%C3%A4higen_Dateiformate.pdf (02.03.2020).

⁴⁷ Vgl. Landesarchiv Thüringen, Empfehlungen zur Anbietung und Übergabe elektronischer Unterlagen an die Thüringischen Staatsarchive. Version 3.3 (2013), https://www.thueringen.de/mam/th1/staatsarchive/behorden/empfehlung_uebergabe_elektronischen_archivguts_v3.3.pdf (02.03.2020).

untersuchten Informationstypen orientiert sich am KaD; die Gruppierung erfolgt nach Text, Rastergrafik, Audio, Video, Strukturierte Daten (Tabellen), Datenbanken, Webarchivierung, Geoinformationssysteme sowie CAD/CAM/Vektorgrafiken. Für eine bessere Vergleichbarkeit musste teilweise abstrahiert und verallgemeinert werden, da die Angabe der Granularität in den einzelnen Katalogen recht unterschiedlich ausfällt.⁴⁸ Der Versuch der Vereinheitlichung stößt jedoch an gewisse Grenzen, da die Terminologie nicht immer klar ist: Während manche Einrichtungen Dateiformate angeben, vermerken andere Kompressionsstandards oder Codierungen. Trotz der Bemühung um Vereinheitlichung konnten solche Widersprüche nicht immer aufgelöst werden.

Die Einschätzungen zur Archivfähigkeit der Formate gehen teilweise weit auseinander. Während der KaD beispielweise nur PDF/A als uneingeschränkt und zwei weitere Formate als bedingt geeignet für die Langzeitarchivierung von Textdokumenten einschätzt, sind es in den Augen der britischen National Archives vierzehn. Dabei ist weiterhin zu beachten, dass die National Archives auch Formate zulassen, die die KOST bisher noch keiner Analyse unterzogen hat. Die Varianz ist somit groß. Dennoch gibt es einen Kern an Formaten, der weitgehend Akzeptanz findet: für Texte handelt es sich um PDF/A, für Rasterbilder um TIFF und JPEG2000, für Audio um WAVE, für Video um die Standards MPEG-2 und MPEG-4. Für diese Informationstypen haben fast alle in der Auswertung berücksichtigten Institutionen Empfehlungen ausgesprochen.⁴⁹ Dagegen wurden die Bereiche der Datenbanken, der Webarchivierung, der Geoinformationssysteme und CAD/CAM/Vektorgrafiken nur von einem Teil bewertet und die Tendenzen zu einem allgemein anerkannten Format oder Standard sind hier weit geringer ausgeprägt. Die durchaus unterschiedlichen Einschätzungen verdeutlichen, dass es das eine, von allen akzeptierte Standardformat nicht für alle Bereiche gibt und vermutlich auch nicht geben wird. Dies muss auch nicht nötig sein, denn solange man sich im Austausch über die Anforderungen an Archivierungsfähigkeit befindet, kann man bei der einzelnen Bewertung eines Formats durchaus zu einem anderen Ergebnis kommen, da je nach Institution, deren Beständen und Umfeld die Gewichtung der Kriterien verschieden ausfallen darf. Entscheidend ist, dass eine systematische Evaluierung überhaupt vorgenommen wird. Ob diese am Ende zu einem festen Formatkatalog in Form einer Whitelist und unter Umständen auch einer

⁴⁸ Beispielweise sind die Angaben des britischen National Archives im Gegensatz zum KaD sehr viel feiner und detaillierter, indem einzelne Versionsnummern differenziert werden.

⁴⁹ Etwas überraschend wurde TIFF in der Liste des Landesarchivs Rheinland-Pfalz gar nicht geprüft.

Blacklist mit völlig ungeeigneten und obsoleten Formaten führen muss, ist jeder Institution überlassen, doch erleichtert ein solcher Katalog nicht nur den Austausch mit Kollegen, sondern er fordert dazu auf, sich kontinuierlich mit der Thematik zu beschäftigen und den Katalog aktuell zu halten.

Durch die Whitelists herrscht nun zwar mehr Klarheit darüber, welche Formate sich am besten zur Übernahme oder als Zielformate einer Migration eignen, aber sie geben noch keine Auskunft darüber, welchen Formaten Obsoleszenz droht. Ein Beispiel für eine Blacklist mit bereits obsoleten Formaten wurde nicht gefunden, lediglich Zusammenstellungen unbrauchbarer Speichermedien.⁵⁰ Der Eintrag in der englischsprachigen Wikipedia mit einer Liste von Dateiformaten führt zumindest acht aus einigen hundert Formaten auf, die obsolet sein sollen.⁵¹ Dies kann als Anzeichen dafür gewertet werden, dass nur wenige Formate bisher von Obsoleszenz betroffen sind. Diese Sicht wird auch vom Informatiker David Stuart Holmes Rosenthal in verschiedenen Blogbeiträgen seit 2007 vertreten.⁵² Vermutlich ist weniger Obsoleszenz, sondern stärker der Wildwuchs in digitalen Magazinen das Problem, da das Überwachen einer Unzahl an Formaten kaum machbar erscheint.⁵³

V. Formaterkennung und -validierung

Bei der Formaterkennung und -validierung handelt es sich um einen weiteren grundlegenden Baustein der digitalen Bestandserhaltung. Nicht immer steckt in einem Format auch das drin, was draufsteht. Allein die Eingabe des Dateinamens „.pdf“ führt bei PRONOM zu 30 Treffern. Die Formaterkennung identifiziert ein Format und gibt Aufschluss über dessen Spezifikationen, die allein durch den Blick auf die Dateiendung nicht erschlossen werden kann. Die Validierung hingegen überprüft, ob das

⁵⁰ Siehe bspw. die „Chamber of Horrors“ der MIT Libraries: <https://dpworkshop.org/dpm-eng/oldmedia/chamber.html> (03.03.2020); oder die Privatsammlung eines britischen Bibliothekars: <https://obsoletemedia.org/> (03.03.2020).

⁵¹ Dabei handelt es sich ausschließlich um OpenOffice.org XML-Formate (STW, SXW, SXD, SXM, STI, SXI, STC, SXC), vgl. https://en.wikipedia.org/wiki/List_of_file_formats (03.03.2020). Der Wikipedia-Artikel zu OpenOffice.org XML (https://en.wikipedia.org/wiki/OpenOffice.org_XML, 03.03.2020) wiederum gibt an, dass es noch Software gäbe, die diese Formate darstellen könne. In den entsprechenden Einträgen in PRONOM sind keine Informationen zur Zugänglichkeit von Software hinterlegt, in der Datenbank der LOC sind diese Formate nicht gelistet.

⁵² Vgl. u.a. Rosenthal, David Stuart Holmes, Format Obsolescence: Scenarios (29.04.2007), <https://blog.dshr.org/2007/05/> (03.03.2020); Ders., Format Obsolescence: the Prostate Cancer of Preservation (07.05.2007), <https://blog.dshr.org/2007/05/format-obsolescence-prostate-cancer-of.html> (03.03.2020); Ders., The Half-Life of Digital Formats (24.11.2010), <https://blog.dshr.org/2010/11/half-life-of-digital-formats.html> (03.03.2020). Auch im Digital Preservation Handbook der DPC (2. Aufl. 2015), <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards> (03.03.2020), wurde festgestellt, dass die Gefahr der Obsoleszenz im Gegensatz zu vor zehn Jahren heute sehr viel geringer eingeschätzt wird.

⁵³ Vgl. ebd.

übermittelte Format tatsächlich alle Kriterien dieses Formats erfüllt. Kenntnisse über die Spezifikationen sind im weiteren Bestandserhaltungsprozess unerlässlich, da beispielsweise im Rahmen einer Migration die bekannten Spezifikationen interpretiert und in ein neues Format umgewandelt werden müssen. Überhaupt benötigt die Planung von Erhaltungsschritten Kenntnis darüber, welche Formate im digitalen Magazin lagern.⁵⁴ Voraussetzung hierfür ist jedoch, dass die vorliegende Datei tatsächlich in Übereinstimmung aller Spezifikationen aufgebaut ist – die Formatidentifizierung allein ist noch keine Garantie dafür.⁵⁵

Die Identifizierung eines Formats kann anhand unterschiedlicher Merkmale erfolgen. Wie das Beispiel „.pdf“ zeigt, handelt es sich bei der Erkennung anhand der Dateinamenserweiterung um eine oberflächliche Variante. Eine JPEG-Datei kann wiederum sowohl unter der Dateiendung „.jpg“, „.jpe“ als auch „.jpeg“ firmieren. Auch wenn die Benennung eines Formats anhand der Dateinamenserweiterung ein geringer Aufwand ist, stellt sie keine belastbaren Informationen zur Spezifikation zur Verfügung. Darüber hinaus gibt sie keine Hinweise auf die Version des Formats, auch kann eine Dateiendung einfach geändert werden, wovon das eigentliche Format unberührt bleibt.⁵⁶ Eine weitere Möglichkeit ist die Identifizierung über den MIME-Type. Hierbei handelt es sich um eine Klassifizierung von Dateien insbesondere für die Kommunikation zwischen Webserver und Browser. Der MIME-Type gibt an, um welchen Medientypen es sich bei der Datei handelt, also die Dateigruppe wie zum Beispiel „text“ oder „image“. Dieser stellt er einen Subtyp nach, der die Art der Datei angibt, z.B. „text/pdf“ oder „image/jpeg“. Jedoch ist die Erkennung anhand des MIME-Type recht ungenau, Formatversionen werden nicht abgebildet. So ist eine genauere Angabe wie die pdf-Variante PDF/A nicht möglich, es kann allein nur PDF erkannt werden.⁵⁷

⁵⁴ Vgl. Bachmann, Steffen/Ernst, Katharina, Formaterkennung. Ziele, Herausforderungen, Lösungsansätze, in: Manke, Matthias (Hrsg.), Auf dem Weg zum digitalen Archiv. Stand und Perspektiven von Projekten zur Archivierung digitaler Unterlagen. 15. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 2. und 3. März 2011 in Schwerin, Schwerin 2012, S. 69–73, hier S. 69.

⁵⁵ Vgl. Funk, Stefan E./Neubauer, Matthias, 7.4 Formatcharakterisierung, in: Neuroth, Heike/Oßwald, Achim/Scheffel, Regine/Huth, Karsten (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010), <http://nbn-resolving.de/urn:nbn:de:0008-2010071949> (28.01.2020), Kap. 7:13–Kap. 7:15.

⁵⁶ Vgl. ebd., Kap. 7:13f.; Kortyla, Stephanie, Digitale Archivierung. (Zu-)Fälle bei der Formatidentifizierung am Beispiel von PRONOM und DROID, in: Sächsisches Archivblatt 2017, H. 1, S. 20f., hier S. 20; Bachmann/Ernst, Formaterkennung, S. 71.

⁵⁷ Eine Liste der MIME-Typen wird von Internet Assigned Numbers Authority (IANA) verwaltet: <https://www.iana.org/assignments/media-types/media-types.xhtml> (04.03.2020). Vgl. https://en.wikipedia.org/wiki/Media_type (04.03.2020).

Am zuverlässigsten, wenn auch nicht vollumfänglich, gilt die Erkennung mittels der Magic Number bzw. Signature. Dabei handelt es sich um die Identifizierung anhand der jedem einzelnen Format spezifischen Bitsequenzen (Signatures). Bestimmte Bitfolgen (auch Magic Numbers genannt) können an unterschiedlichen Stellen im Bitstrom der Datei stehen – oftmals am Dateianfang oder -ende.⁵⁸ Bei der Identifizierung wird der gesamte Bitstrom oder nur Teile davon nach solchen Mustern abgetastet, die dann einem Format zugeordnet werden können.⁵⁹ Die PRONOM Datenbank der britischen National Archives bindet die Magic Numbers in die selbst erstellten Signatures ein, die eine Unterscheidung von beispielweise PDF/A-1a und PDF/A-1b möglich machen. So werden in PRONOM oftmals nicht nur eine, sondern mehrere Bitfolgen und ihre Position für ein Format vermerkt. Zur Identifizierung von PDF/A-1b ist sowohl eine kürzere Bitfolge am Dateianfang als auch eine längere an einer variablen Position innerhalb der Datei angegeben. Dies ermöglicht die Unterscheidung der Versionen.⁶⁰

Aber auch die Formaterkennung mittels der Signature hat ihre Grenzen: Fehlt einer Datei die für PDF/A-1b charakteristische Bitsequenz, wird diese nicht als PDF/A-1b erkannt. Enthält jedoch eine Textdatei die entsprechende Bitsequenz, wird diese fälschlicherweise als PDF/A-1b identifiziert. Aufgrund solcher Unschärfen sollte auf die Identifizierung immer eine Validierung folgen. Denn auch wenn die charakteristische Bitsequenz vorhanden ist, doch andere Spezifikationen des PDF/A-Standards nicht erfüllt werden, kann es sich nicht um eine PDF/A-1b Datei handeln. Die Validierung überprüft somit, ob das Format jede einzelne Spezifikation erfüllt, denn nur dann ist die Datei valide. Zwingende Voraussetzung dafür ist aber eine erfolgte Formaterkennung, denn nur durch diese ist überhaupt klar, welche Spezifikationen überprüft werden müssen. Die Formaterkennung gibt also den Maßstab der Validierung vor.⁶¹ Ausgehend von diesen Ergebnissen kann die richtige Darstellungssoftware gewählt werden.

⁵⁸ Vgl. Kortyla, Stephanie/Treu, Christian, Nutzen und Grenzen der Formatidentifizierung, in: Informationswissenschaft: Theorie, Methode und Praxis 5 (2018), S. 83–95, hier S. 86.

⁵⁹ Vgl. Funk/Neubauer, 7.4 Formatcharakterisierung, Kap. 7:14; Kortyla, Digitale Archivierung, S. 20; Röthlisberger-Jourdan, Claire, Sinn und Zweck der Formaterkennung und deren Grenzen. nestor-Praktikertag 2017, Kiel, 28.06.2017, https://kost-ceco.ch/cms/dl/6d2aa34e60f26b007d3eb3f0f86b061d/Formaterkennung_v1.0.ppsx (06.03.2020).

⁶⁰ Vgl. <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1100&strPageToDisplay=signatures> (06.03.2020). S.a. die Erläuterung der nestor-AG „Formaterkennung“ zur Erstellung von PRONOM Formatsignaturen: <https://wiki.dnb.de/pages/viewpage.action?pageId=115213928> (06.03.2020).

⁶¹ Vgl. Röthlisberger-Jourdan, Claire, Formaterkennung und Formatvalidierung. Theorie und Praxis, in: Keitel, Christian (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises

Sowohl zur Formaterkennung als auch zur -validierung steht mittlerweile eine ganze Reihe von Tools zur Verfügung, die eine automatisierte Analyse der Dateien ermöglichen. Bisher ist jedoch noch kein Tool im flächendeckenden Einsatz, das beide Aufgaben umfassend in sich vereint. Dies hängt auch damit zusammen, dass es zwar Erkennungsprogramme gibt, die (fast) alle Dateiformate identifizieren können, es aber an einem Validator fehlt, der die Erfüllung der Spezifikation aller Formate abprüfen könnte. Bei vielen Validatoren handelt es sich um Spezialtools, die sich auf bestimmte Formate wie PDF spezialisiert haben, in der Praxis fehlt es an Validatoren für viele andere, insbesondere seltene Formate.⁶²

Bei DROID handelt es sich um eines der weitverbreitetsten Open-Source-Erkennungstools, das von den britischen National Archives entwickelt wurde und – wenig überraschend – als Grundlage die Signature Files von PRONOM verwendet.⁶³ Neben der Erkennung anhand des Signature File und der Angabe der PRONOM PUID wirft DROID zusätzlich den MIME-Type aus. Andere gängige Tools wie Siegfried⁶⁴ und FIDO⁶⁵ nutzen ebenfalls PRONOM als Basis, wohingegen TIKA⁶⁶ sich auf eine eigene Datenbank stützt, die insbesondere auf den MIME-Typen aufbaut und in der Granularität mit den PRONOM-basierten Tools nicht mithalten kann.⁶⁷

Die DROID Signature File wird laufend überarbeitet, zuletzt erschien im Januar 2020 Version 96. In dieser Weiterentwicklung liegt gleichwohl ein Problem von DROID, wie Michelle Lindlar und Yvonne Tunnat zeigen konnten.⁶⁸ Durch die Überarbeitungen, detaillierte Versionierungen und neu aufkommenden Formate kam es auch zu Veränderungen der PUIDs, zum Beispiel wurde für TIFF ab Version 51 der

„Archivierung von Unterlagen aus Digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“, Stuttgart 2013, S. 193–210, hier S. 196–199. Für weitere Fehlermöglichkeiten in DROID s.a. Kortyla/Treu, Nutzen und Grenzen, S. 87–90.

⁶² Vgl. ebd., S. 85.

⁶³ DROID (Digital Record Object Identification), <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> (23.03.2020). Es wird auch in Ingestlist verwendet.

⁶⁴ Hierbei handelt es sich um eine Eigenentwicklung des amerikanischen Archivars Richard Lehane der New South Wales State Archives, <https://www.itforarchivists.com/> (06.03.2020).

⁶⁵ Format Identification for Digital Objects (FIDO) wird von der OPF als Open-Source-Software zur Verfügung gestellt, vgl. <https://openpreservation.org/technology/products/fido/> (06.03.2020).

⁶⁶ Tika wird von der Apache Software Foundation betrieben und zum freien Download bereitgestellt: <http://tika.apache.org/> (06.03.2020).

⁶⁷ Vgl. Röthlisberger-Jourdan, Formaterkennung und Formatvalidierung, S. 201–205.

⁶⁸ Vgl. Lindlar, Michelle/Tunnat, Yvonne, Time-Travel with PRONOM. The 4th Dimension of DROID, in: Proceedings of iPRES 2018 15th International Conference on Digital Preservation (2018), https://zenodo.org/record/3517767/files/iPres2018_paper_85_final.pdf?download=1 (27.01.2020).

Signature File eine andere PUID verwendet.⁶⁹ Aus diesem Grund führt die Erkennung der gleichen Testdateien mit älteren Signature Files teilweise zu anderen Ergebnissen als mit der aktuellen. Insbesondere die Versionierungen von PDF riefen Probleme bei der Identifizierung zwischen Version 38 und 48 in den Jahren 2010 und 2011 hervor. Für PDF 1.3 sprechen Lindlar und Tunnat zwischen Version 38 (August 2010) und Version 74 (März 2014) sogar von einem „dark age“⁷⁰. Dies stellt die Archive vor ein Problem, da die Formaterkennung in den meisten Fällen nur beim Ingest oder im Rahmen einer Migration vorgenommen wird. In den Zeiträumen dazwischen, die Jahre und viele neue Signature Files bedeuten können, bleiben die Formate in den digitalen Magazinen unberührt und somit nach dem aktuellen Stand teilweise unerkannt oder falsch erkannt. Die Konsequenz daraus müsste eine Re-Identifizierung dieser Formate sein. PRONOM veröffentlicht im Schnitt zwei- bis dreimal im Jahr eine neue Version der Signature File. Nun muss nicht sofort das ganze digitale Magazin erneut mit DROID gescannt werden, sondern sollten zunächst die auf der PRONOM-Website in den Release Notes dokumentierten Aktualisierungen und Neueintragen begutachtet werden.⁷¹ Veränderungen in der Signature File beispielweise zu sehr exotischen Videoformaten müssen nicht unbedingt umgehend eine Re-Identifizierung entsprechender Videodateien in Gang setzen, eine neue PDF-Version vielleicht hingegen schon. Die kontinuierliche Beobachtung der Signature File ermöglicht es, anlassbezogen und punktuell eine Neuerkennung einzelner Formate durchzuführen. Die Entscheidung für einen solchen Schritt bleibt abhängig von der Bewertung der Veränderungen durch den Archivar, aber auch von der Bedeutung und Struktur der betroffenen Archivalien.

Statistiken zu vorhandenen Dateiformaten im digitalen Magazin, welche die Basis für Planung und Umsetzung von Erhaltungsmaßnahmen bilden, sind jedoch nur so aussagekräftig, wie die Erkennung der Formate aktuell ist. Bisher existieren keine Richtlinien, in welchen Abständen dies erfolgen sollte. Sinnvoll scheint es jedoch, sich neben der anlassbezogenen auch auf eine periodische Prüfung zu einigen. Aufgrund der Dynamik wäre eine Neuprüfung alle zehn Jahre ein zu großer Zeitraum, hingegen sind umfassende jährliche Scans personell und organisatorisch wenig realistisch. Als Mittelweg wird hier neben der anlassbezogenen Neuerkennung eine Überprüfung

⁶⁹ In Version 1–4 konnte TIFF nicht erkannt werden, da kein Eintrag in PRONOM vorhanden war. Ab Version 5 gab es hingegen vier unterschiedliche PUIDs für TIFF (fmt/7, fmt/8, fmt/9, fmt/10), die ab Version 51 durch eine PUID (fmt/353; Juli 2011) ersetzt wurden.

⁷⁰ Lindlar/Tunnat, Time-Travel.

⁷¹ Vgl. <https://www.nationalarchives.gov.uk/aboutapps/pronom/release-notes.xml> (06.03.2020).

mindestens der aktuellsten Repräsentationen, wenn nicht des gesamten Magazins alle fünf Jahre vorgeschlagen.⁷²

Zur Validierung von Formaten wird mehrheitlich das Open-Source-Werkzeug JHOVE angewendet, das aus einem Projekt von JSTOR und der Harvard University Library hervorgegangen ist und nun von der OPF betrieben wird. JHOVE ist derzeit in der Lage, siebzehn verschiedene Formate und deren dazugehörigen Versionen zu identifizieren und zu validieren, dabei handelt es sich insbesondere um Text-, Bild und Tonformate.⁷³ Daneben bestehen Spezialtools für einzelne Formate wie BadPeggy für JPEGs oder veraPDF für PDF/A-Dateien.⁷⁴

Den Überblick über die Menge an Tools zu behalten, ist nicht leicht. Abhilfe kann die Community Owned digital Preservation Tool Registry (COPTR) der OPF oder die nestor-Toolbox bieten.⁷⁵ Dabei handelt es sich um Sammlungen, Beschreibungen und Einschätzungen unterschiedlicher Werkzeuge, ihrer Funktion und Einsatzgebiete – nicht nur, aber eben auch für die Formaterkennung und -validierung.

Um der Zersplitterung der Landschaft der Validierungstools entgegen zu wirken, gibt es Ansätze, unter dem Dach eines Tools verschiedene andere zu vereinen. Die KOST hat mit KOST-Val ein Programm zur Validierung von TIFF-, SIARD-, PDF/A-, JP2- und JPEG-Dateien sowie SIPs entwickelt, das unterschiedliche bestehende Tools wie JHOVE, pdfapilot und BadPeggy miteinander verbindet.⁷⁶ Auch bei dem Tool FITS der Harvard Library handelt es sich um ein Verbindungstool, das im Gegensatz zum KOST-Val nicht nur Validierungswerkzeuge versammelt, sondern

⁷² Grundsätzlich sollte eine Überprüfung der aktuellsten Repräsentation ausreichen, jedoch sind auch Migrationen aus älteren Repräsentationen denkbar, weshalb auch für diese eine aktuelles Identifizierungsergebnis vorliegen müsse. S.a. Anm. 72.

⁷³ Vgl. die „Modules“ auf <http://jhove.openpreservation.org/> (06.03.2020). JHOVE ist in Ingestlist eingebunden. Mit JHOVE2 wurde ein Nachfolger entwickelt, der auf einer anderen Softwarearchitektur beruhte. Nachdem die Projektfinanzierung 2011 ausgelaufen ist, wurde seit der letzten Version von 2013 die Weiterentwicklung von JHOVE2 eingestellt. Vgl. <https://coptr.digipres.org/JHOVE2> (06.03.2020).

⁷⁴ Bei veraPDF handelt es sich auch um ein Open-Source-Tool, das mittlerweile von der OPF, PDF Association und Digital Preservation Coalition getragen wird. Bis vor einigen Jahren gab es kein Angebot an Open-Source-Validatoren für PDF/A, vgl. Röthlisberger-Jourdan, Formaterkennung und Formatvalidierung, S. 207.

⁷⁵ Vgl. https://coptr.digipres.org/Main_Page (06.03.2020). COPTR listet aktuell 22 Tools zur Formaterkennung und 43 zur Validierung auf. Für die nestor-Toolbox vgl. <https://wiki.dnb.de/pages/viewpage.action?pageId=134715087> (06.03.2020). Aufgrund der bestehenden Tool Registry COPTR strebt die nestor-AG „Formaterkennung“ jedoch nicht den Aufbau einer eigenen Registry an, sondern unterstützt durch Beiträge COPTR. Die Toolbox hingegen ist aus einer Fortbildungsveranstaltung auf dem Deutschen Archivtag 2015 entstanden und wird seitdem ohne Anspruch auf Vollständigkeit weiter gepflegt. Für weitere, teils veraltete Links zu Tool Registries vgl. DPC, Digital Preservation Handbook, 2. Aufl. 2015, <https://www.dpconline.org/handbook/technical-solutions-and-tools/tools> (08.03.2020).

⁷⁶ Vgl. <https://kost-ceco.ch/cms/kost-val.html> (06.03.2020).

auch solche zur Erkennung. Neben DROID und TIKA wird unter anderem auch JHOVE eingebunden. Die Ergebnisse der einzelnen Werkzeuge werden in eine einheitliche Ausgabeform gebracht und miteinander verglichen.⁷⁷ Yvonne Tunnat und Michelle Lindlar haben jedoch darauf hingewiesen, dass FITS nicht die aktuelle DROID Version verwendet und somit unter Umständen Aussagen auf Grundlage einer veralteten Signature File trifft.⁷⁸

Diese Ansätze und COPTR zeigen, dass der Bedarf an Orientierung und Bündelung nicht nur im Bereich der Archivierungsformate, sondern auch bei den damit verbundenen Werkzeugen groß ist. In Anbetracht unzähliger Tools für noch mehr Formate muss die Entwicklung hin zu einer stärkeren Automatisierung und Verzahnung gehen. Auf Grundlage des Ergebnisses der Formaterkennung sollte automatisiert der passende Validator ausgewählt werden. Zu überlegen wäre, inwieweit Identifizierungstools miteinander verknüpft werden könnten. Asgar Blekinge hat 2012 DROID, FIDO und TIKA einem Test unterzogen und deren Ergebnisse anhand eines Testkorpus zu unterschiedlichen Formaten verglichen.⁷⁹ Sein Ergebnis war, dass TIKA bestimmte Formate zuverlässiger erkennt als die anderen beiden, während eines der anderen beim nächsten Format wiederum zuverlässiger arbeitet als TIKA, manche Formate aber von allen gleich gut oder schlecht erkannt werden. Da die Qualität der Identifizierung somit stark davon abhängt, ob man für das jeweilige Format das richtige Tool verwendet, wäre ein Stufensystem denkbar. In einem ersten Schritt könnte zunächst eine Grobeinteilung anhand der Dateiendung oder des MIME-Types erfolgen. In einem zweiten Schritt sollte ein Formaterkennungstool ausgewählt werden, das sich für diese Formatkategorie (bisher) am besten bewährt hat. Möglich wäre auch die Auswahl eines Tools anhand seiner Fähigkeit, die benötigte Granularität zu bestimmen, die in bestimmten Konstellationen vielleicht nicht immer die feinste sein muss. Ebenso sollten die Bearbeitungszeiten der Tools in die Bewertung einbezogen werden. Unter Umständen ist eine leicht erhöhte Fehlerquote zu akzeptieren, sofern die Erkennung von großen Mengen von Dateien in einem Durchgang in einer angemessenen Zeit erfolgen kann. Anhand der Identifizierung des Formates wird nun das entsprechende Validierungswerkzeug gewählt. Auch hier ist zu testen, welcher der Validatoren der

⁷⁷ Vgl. <https://projects.iq.harvard.edu/fits> (06.03.2020).

⁷⁸ Vgl. Lindlar/Tunnat, Time-Travel.

⁷⁹ Vgl. Blekinge, Asgar, Identification Tools, an Evaluation. The Scape Characterisation Tool Testing Suite (2012), <https://openpreservation.org/blog/2012/02/23/identification-tools-evaluation/> (14.02.2020). Auf Bleskinges Test aufbauend hat auch die KOST einen solchen unternommen, vgl. Röhllisberger-Jourdan, Formaterkennung und Formatvalidierung, S. 201–207.

geeignetste ist.⁸⁰ Dies setzt aufwendige Testverfahren voraus, um genügend belastbare Informationen über die Zuverlässigkeit, Genauigkeit und Geschwindigkeit der Erkennungstools zu erhalten. Auch müssten diese Tests mit angepassten Dateikorpora regelmäßig durchgeführt werden, da sich nicht nur die Formate, sondern auch die Werkzeuge und die ihnen zugrundeliegenden Datenbanken weiterentwickeln. Für viele, besonders für die gängigen Formate, könnten feste Prüfungs- und Toolfolgen definiert werden, die dann automatisiert ablaufen. Auch hierbei handelt es sich um eine Aufgabe, die ein Archiv allein nur in den seltensten Fällen bewältigen kann, sondern für die Archive auf die Kooperation in der internationalen Fachwelt angewiesen ist. Aus diesem Grunde ist das Engagement von nestor und der OPF, die solche Testsuites zum Testen und Evaluieren von Tools bereits zur Verfügung stellen, nur zu begrüßen.⁸¹

VI. Formate in DIMAG – eine kleine Bestandsaufnahme

DIMAG stellt dem Nutzer eine Gesamtstatistik mit Informationen über die Zahl der vorhandenen Objekte, die Gesamtzahl der Datensätze, den Umfang des belegten Speichers, die Anzahl der gelöschten Objekte und die vorhandenen Formate mit der entsprechenden Zahl an Dateien zur Verfügung. Diese Abfrage erfolgt über alle Strukturebenen hinweg. Das Landesarchiv Baden-Württemberg hatte am 8. Januar 2020⁸² 1.663.363 Primärdatenobjekte in DIMAG abgelegt. Diese Primärdatenobjekte bestanden aus 328 unterschiedlichen Formaten, wobei die Zahl nach einer groben Bereinigung auf rund 240 Formate schrumpft.⁸³ Es wurden insgesamt nur 10 unterschiedliche MIME-Types ausgemacht. Hieran wird deutlich, dass eine alleinige Formaterkennung

⁸⁰ Ähnliche Überlegungen für einen Workflow zur Formatverifikation haben Christine Träger und Daniel Wittmann bei einer Tagung des AUdS 2019 vorgestellt, der auf den Präsentationsfolien in groben Zügen nachvollzogen werden kann: Träger, Christine/Wittmann, Daniel, Validierst du noch oder archivierst du schon... Der Bedarf eines Workflow-Managements in der Formatverifikation. 23. Tagung des Arbeitskreises AUdS, Prag, 12. und 13. März 2019, https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/camp-2-f/F_01_Traeger_AUdS_Praesentation_Formatverifikation.pdf (06.03.2020).

⁸¹ Vgl. <https://github.com/openpreserve/format-corpus> (06.03.2020); <https://github.com/nestorFormatGroup/FileTestSuite> (06.03.2020).

⁸² Die dieser Arbeit zugrundeliegende DIMAG-Statistik spiegelt den Stand vom 08.01.2020 wider.

⁸³ Die Statistik basiert dabei auf den Ergebnissen der Formatcharakterisierung beim Ingest insbesondere mit DROID, weshalb die PUID, der Formatname gemäß der PUID und der MIME-Type angegeben werden. Bei der Bereinigung erfolgte grundsätzlich eine Orientierung an den PUIDs, doppelte Einträge und Inkonsistenzen wurden gelöscht. Beispielsweise wurde in den Fällen, in denen eine PUID doppelt auftaucht, aber nur bei einem Eintrag auch der MIME-Type angegeben wurde, nur ein Format und nicht zwei gezählt. Auch wurden unterschiedliche Auszeichnungen derselben Codierung im MIME-Type wie „charset= US-AS“ und „charset= US-ASCII“ nicht doppelt gezählt. Jedoch wurden Einträge mit derselben PUID, die unterschiedliche Codierungen vorwiesen wie „charset=UTF-8“ und „charset=US-ASCII“, als zwei Einträge beibehalten. Insgesamt wurden 216 unterschiedliche PUIDs erkannt.

anhand von MIME-Types nicht ausreichend sein kann. In Anbetracht von mehr als 1.300 Formateinträgen in PRONOM und mehr als 1.500 bestehenden MIME-Types handelt es sich grundsätzlich um ein noch überschaubares Feld vorhandener Formate, das nichtsdestotrotz für die Erarbeitung differenzierter Erhaltungsmaßnahmen schnell unübersichtlich wird. Dazu trägt der Umstand bei, dass auf 207 unterschiedliche Formate jeweils weniger als 50 Dateien fallen, darunter wiederum auf 142 Formate weniger als 10 Dateien.

Mit 223.603 Dateien macht das Bildformat JPEG File Interchange Format Version 1.00 den größten Bestand aus. Ein Blick auf die ersten zehn Plätze der häufigsten Dateiformate in DIMAG zeigt, dass Bilddateien in einem JPEG- oder TIFF-Format und Varianten von Shapefiles mit Abstand die meisten Primärdatenobjekte ausmachen. Zusammen kommen sie auf rund 970.000 Dateien, umfassen somit fast zwei Drittel aller Primärdatenobjekte in DIMAG.⁸⁴ Rund 81.000 Bilder im langzeitarchivtauglichem TIFF-Format stehen insgesamt mehr als 400.000 Dateien im JPEG-Format⁸⁵ gegenüber, das aufgrund seiner Komprimierung eigentlich zur Langzeitarchivierung ungeeignet gilt, von vielen Archiven aber dennoch als Übernahmeformat akzeptiert wird, da es sich um eines der weitverbreitetsten Formate mit vielen Implementierungsmöglichkeiten handelt. Dateien, die schon im verlusthaften JPEG-Format angeboten werden, werden von den meisten Einrichtungen übernommen und derzeit nicht in ein verlustfreies Format konvertiert.⁸⁶ Damit ist jedoch nicht ausgeschlossen, dass in Zukunft JPEG-Dateien migriert werden müssen. Der große Umfang an Shapefiles beruht darauf, dass es sich um den gängigen Industriestandard für Geoinformationssysteme handelt, der grundsätzlich als langzeitarchivierungsfähig gilt.⁸⁷ Ein Großteil der Bestände in DIMAG liegt somit in derzeit als archivierungstauglich angesehenen Formaten vor. Da es an Blacklists fehlt, kann nicht eingeschätzt werden, wie viele potentiell bedrohte Formate im digitalen Magazin schlummern; bereits obsoletere Formate sind aufgrund ihrer Rarität nicht zu erwarten.

⁸⁴ Dahingegen liegen nur ca. 41.000 PDF-Dateien vor, wovon nur rund die Hälfte PDF/A-Dateien sind.

⁸⁵ Hier wurden die Dateien von JPEG File Interchange Format V1.00, V1.01 und V1.02 zusammengekommen.

⁸⁶ Vgl. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000017.shtml> (09.03.2020); <https://kost-ceco.ch/cms/jpeg.html> (09.03.2020).

⁸⁷ Vgl. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml> (09.03.2020). Die KOST schätzt es als langzeitarchivierungstauglich ein, auch wenn in der Schweiz INTERLIS der Standard im Bereich der GIS ist, vgl. <https://kost-ceco.ch/cms/shapefile.html> (09.03.2020).

Auffällig ist, dass bei rund 160.000 Dateien, also ca. 10 % des Gesamtbestandes, der Eintrag „Das Format konnte nicht identifiziert werden!“ erscheint. Ein genauer Blick zeigt, dass bei 78.220 Dateien weder eine PUID noch ein MIME-Type ermittelt werden konnte, wohingegen bei ca. 3.000 mindestens eine PUID und bei ca. 79.000 ein MIME-Type angegeben ist. Dies kann mehrere Gründe haben: Zum einen könnte ein Eintrag zum Format in PRONOM noch nicht vorgelegen haben oder zum Zeitpunkt der Erkennung war zwar schon eine PUID und Signature angelegt, jedoch noch keine menschenlesbare Beschreibung, weshalb zwar die PUID, aber nicht der Formatname erkannt wurde. Zum anderen kann es auch beim Identifizierungsvorgang aufgrund der Dateigröße, Erkennungstiefe o.ä. durch das Erkennungstool zu einer Fehlermeldung gekommen sein.⁸⁸ Dennoch bleiben knapp 78.000 Dateien zurück, die bisher nicht identifiziert wurden. Dies ist in Anbetracht des Umstandes, dass mit der neuen DIMAG-Softwareversion von 2016 der komplette Datenbestand neu eingelesen und somit neu identifiziert wurde, ein hoher Wert. Eine Re-Identifizierung dieser Dateien wäre angebracht, da immer zu bedenken ist, dass veraltete und fehlerhafte Identifizierungen die Statistik verfälschen.

Für Rückschlüsse aus der DIMAG-Statistik auf notwendige Bestandserhaltungsmaßnahmen ist es problematisch, dass die Statistik keine Differenzierung nach Repräsentationen erlaubt. Die bisherigen statistischen Angaben zu den Formaten gehen über alle Repräsentationsstufen hinweg. Für die Bestandserhaltungsplanung ist dies jedoch nicht zielführend, da im Normalfall nur die Formate der Primärdatenobjekte der zuletzt angelegten Repräsentation von Interesse sind.⁸⁹ Sofern schon eine Repräsentation 2 besteht, können bezüglich der Bestandserhaltung die Dateiformate der Repräsentation 1 außer Acht gelassen werden, da diese schon in ein besser geeignetes Format migriert wurden. Es müsste in DIMAG also eine Abfrage programmiert werden, die jene Repräsentation mit der höchsten Nummer im Signaturanteil oder alternativ mit dem jüngsten Erstelldatum in den Metadaten unterhalb eines Informationsobjekts erfasst. Von der Repräsentation wiederum kann auf die Unterobjekte, also auf die Primärdatenobjekte geschlossen und die entsprechenden Informationen zum

⁸⁸ Hinweise von Boris Kraut, Landesarchiv Baden-Württemberg, E-Mail vom 20.02.2020.

⁸⁹ Es ist durchaus denkbar, dass bedingt durch den technischen Fortschritt die Repräsentation 3 aus der Migration der Dateien von Repräsentation 1 statt Repräsentation 2 erfolgt, da durch neue Formate und Tools sich ein besseres Migrationsergebnis aus den Übernahmedateien erzielen lässt als aus jenen, die schon Ergebnis einer Migration sind. Möglich wäre ein solches Vorgehen auch bei Veränderungen in den signifikanten Eigenschaften. Dennoch ist nach heutigem Stand anzunehmen, dass Migrationen von der jeweils letzten Repräsentation ausgehen werden.

Dateiformat erhoben werden. Des Weiteren wäre es sinnvoll, jene Dateiobjekte auflisten zu können, bei denen bisher kein Format erkannt werden konnte oder deren Formaterkennung unsicher ist – erweitert um die Möglichkeit, auch dies auf die aktuellste Repräsentation begrenzen zu können.

Es bleibt darüber hinaus zu bedenken, dass nicht alle Primärdatenobjekte eines digitalen Magazins tatsächlich erhalten werden müssen. In DIMAG befinden sich beispielsweise über 15.000 Dateien des Formats Binary Property List. Dabei handelt es sich um ein Format zur strukturierten Datenspeicherung innerhalb eines Betriebssystems, das auf Apples iPhones verwendet wird.⁹⁰ Die kompletten Datenbestände von iPhones lassen sich als Beweismittel in den Ermittlungsakten der Staatsanwaltschaften finden. Bei der Übernahme der Ermittlungsakten finden solche Daten ihren Weg in DIMAG, wobei diskutabel ist, ob sie bei der Erstellung einer neuen Repräsentation migriert werden müssen. Es ist aus archivalischer Sicht vorstellbar, dass in vielen Fällen nicht der komplette Inhalt eines Mobiltelefons erhalten werden muss, da die relevanten Informationen, die hier gefunden wurden, sich an anderer Stelle in den Ermittlungsakten niedergeschlagen haben. Zumindest ist eine Priorisierung von Erhaltungsmaßnahmen anhand des Informationsgehalts und der Bedeutung einer Archivale bzw. eines Primärdatenobjektes sinnvoll. Am Ende muss wohl nicht jedes Dateiformat erhalten werden, die Entscheidung gegen die Erhaltung aber protokolliert und nachvollziehbar gemacht werden.

VII. Das Datenmodell der Preservation Action Registries (PAR)

Zur Stärkung der internationalen Vernetzung im Bereich der digitalen Bestandserhaltung wurde 2018 ein Projekt zum Austausch von Erhaltungsmaßnahmen zwischen Anbietern von Archivspeichersystemen und Vertretern der britischen digitalen Forschungsinfrastruktur angestoßen. Getragen durch das JISC⁹¹ und ihrem Research Data Shared Service (RDSS) Pilotprogramm, haben Arkivum, Artefactual und Preservica

⁹⁰ Vgl. <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1789&strPageToDisplay=summary> (10.03.2020); Karaiskos, Christos, Understanding Apple's Binary Property List Format (27.02.2018), <https://medium.com/@karaiskc/understanding-apples-binary-property-list-format-281e6da00dbd> (10.03.2020).

⁹¹ Bei dem Joint Information Systems Committee handelt es sich um eine gemeinnützige Organisation, die britische Hochschulen und Weiterbildungseinrichtungen mit digitalen Dienstleitungen und Lösungen in Forschung und Lehre unterstützt. Das RDSS stellt einen Versuch dar, eine gemeinsame Forschungsdateninfrastruktur zur Verfügung zu stellen.

unterstützt durch die OPF ein Konzept zu sogenannten Preservation Action Registries (PAR) entwickelt.⁹²

Ziel des Projektes ist, bestehende Register von Bestandserhaltungsmaßnahmen zu vereinheitlichen, miteinander zu vernetzen und für Dritte zu öffnen. Die Entwickler von PAR wollen Erhaltungsmaßnahmen nicht fachlich bewerten und Empfehlungen aussprechen (dies überlassen sie der Community), sondern die technischen Voraussetzungen für eine Verbindung verschiedener Register schaffen und diese in maschinenlesbare Regelsätze verwandeln. Sowohl Preservica als auch Artefactual⁹³ pflegen bereits eigene Register, welche Tools nach welchen Regeln für welche Maßnahmen angewandt werden sollen, doch können diese Regelsätze derzeit nicht auf einer technischen Ebene miteinander ausgetauscht werden. Eine Interoperabilität der Register in einer Open-Source-Umgebung mit entsprechenden Schnittstellen soll ermöglichen, dass auch andere Interessenten von dem gesammelten Wissen profitieren und mit ihren „best practice“-Erfahrungen anreichern können. Zu Beginn konzentriert sich das Projekt auf den Austausch zwischen Preservica und Archivemata, PAR soll aber anschlussfähig für weitere Speichersysteme sein.

Im Zentrum des Datenmodells (PAR Conceptual Model; Anlage 2) steht die Entität „Preservation Action“, somit Erhaltungsmaßnahmen wie Checksummen erstellen oder migrieren. Sie wird mit einem „Tool“⁹⁴ auf ein „Object“ oder eine „File“ angewendet. Die Maßnahme an sich und das Werkzeug zur Durchführung sind getrennte Entitäten. Die Preservation Action kann durch Parameter wie einem Kompressionsgrad begrenzt werden. Jede Preservation Action lässt sich einem „Preservation Action Type“ zuordnen, also einer Klassifikation wie Formaterkennung oder Migration. Zur einheitlichen Bezeichnung der Preservation Action Types wird auf den PREMIS

⁹² Für das Folgende vgl. Addis, Matthew/Simpson, Justin/Tilbury, Jonathan/O’Sullivan, Jack/Stokes, Paul, Digital Preservation Interoperability through Preservation Actions Registries (2018), <https://doi.org/10.6084/m9.figshare.6628418.v1> (27.01.2020); <https://openpreservation.org/news/arkivum-artefactual-the-open-preservation-foundation-and-preservica-collaborate-on-new-jisc-initiative-for-sharing-preservation-action-best-practice/> (11.03.2020); Tilbury, Jon, A New Era in Collaboration in Digital Preservation Research (29.06.2018), <https://www.dpconline.org/blog/a-new-era-in-collaboration-in-digital-preservation-research> (11.03.2020).

⁹³ Preservica: Linked Data Registry (LDR), vgl. Braud, Maité/Sinclair, Pauline/Sharpe, Robert, Linked Data Registry. A New Approach to Technical Registries; Archives and Cultural Industries, 11th–15th October 2014, <https://pdfs.semanticscholar.org/98bd/5a5ac26c40438b8af978021bdca8123f1988.pdf> (11.03.2020); Artefactual bzw. Archivemata: Format Policy Register (FPR), <https://www.archivemata.org/en/docs/archivemata-1.10/user-manual/preservation/preservation-planning/> (11.03.2020).

⁹⁴ Für die Bestimmung der Tools ist der Verweis auf Register wie COPTR möglich; hierfür wären eindeutige Identifier in COPTR vorteilhaft. Denkbar wäre auch, die Tools wiederum mit PREMIS als Agent zu beschreiben, da hierunter auch Software in Verbindung mit einem Event fällt.

eventType und das von der LOC empfohlene Vokabular zurückgegriffen.⁹⁵ Gegenstand einer Preservation Action ist entweder eine einzelne Datei oder ein Objekt, das aus mehreren Dateien besteht. Beide können sowohl Input und damit Auslöser, als auch Output und dadurch ein Ergebnis einer Preservation Action sein. Eine Worddatei wäre Auslöser eines Migrationsvorganges an dessen Ende eine weitere Datei, nun als PDF/A entsteht. Jede Datei und jedes Objekt hat bestimmte Eigenschaften („Properties“), die wiederum Gegenstand und Ergebnis einer Preservation Action sein können. Die Checksumme einer Datei wäre eine solche Eigenschaft, die zunächst als Input wirkt, während die durch ein Tool neu generierte Checksumme im Rahmen einer Checksummenprüfung als Output wiederum eine Eigenschaft der Datei darstellt. Beim „File Format“ handelt es sich um eine der wichtigsten Eigenschaften einer Datei und das Zentrum der meisten Preservation Actions, weshalb sie als eigene Entität aufgelistet wird.⁹⁶ Ein Set an Eigenschaften eines Objektes oder einer Datei bildet wiederum ein „Profile“. Hierbei kann es sich um verschiedene kombinierte Metadaten handeln, bei einer Datei beispielsweise um die Größe zusammen mit der Checksumme und dem Format. Es können in diesen Containern jedoch auch Metadaten abgelegt werden, die nicht durch Allgmeinstandards abgedeckt sind.⁹⁷ Die letzte Entität bilden die „Business Rules“, bei denen es sich um eine Sammlung von Aussagen, Hinweisen und Vorschlägen aus der Praxis handelt.⁹⁸ Es sind keine festen Regeln und Vorgaben, sondern Erfahrungswerte der Nutzergemeinde, wie beispielsweise „Tool A ist für das Format X dem Tool B vorzuziehen“ oder „Die Preservation Action A wende ich nur auf Format Y an“. Zusammenfassend ist der Kern des Modells eine Preservation Action, die durch einen Input ausgelöst wird und einen Output als Ergebnis hat. Sie wird mit einem oder mehreren Tools durchgeführt und von einer Business Rule geleitet.

Die Praxis stellen sich die Entwickler von PAR wie folgt vor: Archivar A hinterlegt in seinem Bestandserhaltungssystem die Regel zur Durchführung einer Bestandserhaltungsmaßnahme, dass in seinem Archiv die Metadaten für Quicktime Movie Dateien mit dem Metadatenstandard EBUCore beschrieben werden. Zur Extraktion der Metadaten im EBUCore-XML-Format soll das Tool Mediainfo V 18.03

⁹⁵ Vgl. <http://id.loc.gov/vocabulary/preservation/eventType.html> (11.03.2020).

⁹⁶ Für die eindeutige Zuordnung eines Formats kann wiederum auf die PUIDs von PRONOM zurückgegriffen werden, andere Register wären aber auch denkbar.

⁹⁷ Möglich soll dies u.a. durch die PREMIS objectCharacteristicsExtension sein.

⁹⁸ Es bleibt abzuwarten, ob in einer neuen Version von PREMIS Business Rules als weitere Entität eingebunden werden. Bisher wurde davon Abstand genommen, da es sich nicht um tatsächliche Eigenschaften eines Objektes handele, vgl. PREMIS Editorial Committee, Data Dictionary, S. 33.

verwendet werden. Diese Information exportiert Archivar A in einem vorgegeben PAR Schema zu PAR (Anlage 3 & 4). Hierbei wird sowohl die Preservation Action inklusive Preservation Action Type, Tool, In- und Output beschrieben, als auch die Business Rule einem Preservation Action Type zugeordnet und mit weiteren Hinweisen versehen. Durch die inhaltliche Beschreibung der Business Rules können Priorisierungen von Maßnahmen oder Tools empfohlen werden, indem in einer Business Rule beispielsweise vermerkt wird, dass allgemein zur Formatidentifizierung vorzugsweise DROID verwendet werden sollte, FIDO nur die zweite Wahl darstelle. Die Preservation Action und Business Rule sind mit einer eindeutigen ID innerhalb PAR ausgestattet, wodurch wiederum Querverweise möglich sind. Nun konsultiert Archivarin B aus einem anderen Archiv regelmäßig die neuesten Einträge in der PAR Preservation Action Liste und stößt dabei auf den Eintrag von A. Dieser erscheint ihr sinnvoll, weshalb sie ihn über eine Schnittstelle aus PAR inklusive aller notwendigen Informationen zu Business Rule, Preservation Action, Tool und der PAR-ID in die Datenbank ihres eigenen Bestandserhaltungssystems übernimmt.⁹⁹

Ein Proof of Concept konnte für PAR bereits durchgeführt werden, derzeit sind die Entwickler bemüht, weitere Praxisbeispiele aus der Bestandserhaltungscommunity und weitere Register einzupflegen. Eine Betaversion befindet sich in Vorbereitung. Jedoch läuft die Anfangsfinanzierung durch die JISC aus und die langfristige Weiterfinanzierung scheint noch nicht abschließend geklärt.¹⁰⁰ Es wäre umso misslicher, wenn PAR ein ähnliches Schicksal wie andere vielversprechende, aber am Ende nur kurzlebige Projekte erleiden würde, da es sich um einen fruchtbaren Ansatz handelt, eine Schneise in den Wald von allgemeinen und spezialisierten Registern, Tools und einer großen Menge an Erhaltungsmaßnahmen zu schlagen. Durch die Schnittstellen wäre PAR auch in DIMAG nutzbar.¹⁰¹ Voraussetzung dafür ist jedoch, dass in DIMAG eine Bestandserhaltungsfunktion angelegt ist, die auf einem vergleichbaren

⁹⁹ Vgl. Präsentation der OPF: „Preservation Action Registries Workshop“, iPres Event Workshop, September 2019, <http://parcore.org/presentations/PAR%20Presentation%20iPres%202019.pdf> (12.03.2020). Die Informationen dieser Präsentation insbesondere in Bezug auf die Schnittstellen, IDs, aber auch Anschaulichkeit gehen weit über Addis u.a., Digital Preservation Interoperability, hinaus. Für ein Demo zur Übertragung eines FPR-Eintrages aus Archivematica in PAR vgl. <https://youtu.be/ELjbgksjBmU> (12.03.2020).

¹⁰⁰ Vgl. Präsentation der OPF: „Preservation Action Registries Workshop“, DPC/JISC Workshop, 08.07.2019, <http://parcore.org/presentations/PAR%20Presentation%20York%202019.pdf> (12.03.2020). Die Einbindung der JISC und der OPF sowie die Beteiligung der kommerziellen Firmen geben grundsätzlich Anlass zur Hoffnung, dass eine dauerhafte Finanzierung und Institutionalisierung erreicht werden kann.

¹⁰¹ Für nähere Informationen zu den Schnittstellen vgl. <https://preview.preservica.com/Registry/par/documentation.html> (12.03.2020); <https://github.com/artefactual-labs/rdss-par/wiki> (12.03.2020).

Preservation Action Register beruht, wie es bei Preservica und Archivemata der Fall ist. Dies könnte im Rahmen des von der Digitalen Archivierung Nord (DAN) als DIMAG-Entwicklungspartner konzipierten DIMAG-Bestandserhaltungsmoduls (BEM) umgesetzt werden. Das BEM ist als erweiterbares Framework für Bestandserhaltungsmaßnahmen konzipiert, das im ersten Schritt weitgehend automatisierte Migrationen in das PDF/A-Format umsetzen soll. Für die drei Hauptfunktionen der Formaterkennung, -migration und -validierung mit dem Zielformat PDF/A sind im BEM Format Policies hinterlegt, wobei insbesondere die Auswahl der geeigneten Werkzeuge bei Migration und Validierung den Anwendern Freiräume für individuelle Lösungen lassen und mithilfe des Camunda Modeler neue Policies hinterlegt werden können.¹⁰² Bei diesen Format Policies handelt es sich um Preservation Actions und Business Rules im Sinne von PAR, die von den Anwendern an die eigenen Voraussetzungen und Zwecke angepasst werden können. Durch den Anschluss an PAR könnten Arbeit und Zeit bei der Definition von Erhaltungsmaßnahmen gespart werden, da mindestens ein Grundstock aus PAR importiert werden könnte – vorausgesetzt PAR geht in naher Zukunft in den Produktivbetrieb über. Darüber hinaus bietet es jedem DIMAG-Partner die Möglichkeit, für sein Archiv die jeweils passenden Preservation Actions auszuwählen. Insbesondere für die kleineren Archive im DIMAG-Verbund könnte dies eine große Entlastung, Vereinfachung und somit gleichzeitig Motivation für die digitale Bestandserhaltung bedeuten.

VIII. Protokollierung von Erhaltungsmaßnahmen in DIMAG

Um die Nachvollziehbarkeit jeder Bestandserhaltungsmaßnahme durch spätere Archivre und Nutzer, das Erkennen von Fehlern auch mit zeitlichem Abstand und die Authentizität eines digitalen Archivs zu gewährleisten, ist eine verlässliche Protokollierung aller Handlungen, die an einem Objekt vorgenommen werden, unverzichtbar. Die Protokollierung in DIMAG erfolgt derzeit über Protokollobjekte zu solchen Objekten, die Unterobjekte haben können, also zu Struktur-, Informations- und

¹⁰² Vgl. die Präsentation von Tobegen, Michael, Entwicklung eines DIMAG-Bestandserhaltungsmoduls. Werkstattbericht, in: Naumann, Kai/Dahlmanns, Tim (Hrsg.), Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen. Tagungsdokumentation. Workshop des KLA-Ausschusses Digitale Archive, Bundesarchiv Koblenz, 7./8. November 2018, https://www.bundesarchiv.de/DE/Content/Downloads/KLA/tagungsdokumentation-bestandserhaltung-dig-unterlagen.pdf?__blob=publicationFile (24.02.2020); BEM-Benutzerhandbuch, Version vom 23.05.2019 [DIMAG-Wiki, intern], https://dimag-wiki.la-bw.de/xwiki/bin/download/DIMAG-Wiki+Startbereich/Bestandserhaltungsmodul/BEM_Benutzerhandbuch_2019-05-23.pdf (22.03.2020), S. 11–15.

Repräsentationsobjekten.¹⁰³ Da Primärdatenobjekte die unterste Objektebene darstellen, können unter ihnen keine Protokollobjekte abgelegt werden. Veränderungen an einem Primärdatenobjekt sind somit im Protokoll zur dazugehörigen Repräsentation zu finden. Derzeit wird in diesem Protokoll über das Primärdatenobjekt automatisiert das Erstellen, Statusänderungen inklusive Löschen sowie das Verschieben eines Primärdatenobjektes vermerkt. Eine automatisierte Protokollierung von Migrationsvorgängen besteht noch nicht. Auch werden andere Bestandserhaltungsmaßnahmen wie die Formaterkennung, -validierung oder das Generieren und Prüfen von Checksummen in diesen Protokollen nicht festgehalten. Solche Informationen sind teilweise in den Metadaten des Primärdatenobjekts zu finden.

Es ist fraglich, ob die bisherige Praxis ausreichend ist. Die Informationssicherung in den Metadaten unter anderem mit der Angabe des Formats nach PRONOM inklusive PUID und verwendeter Signature File, die Angabe des MIME-Types, das Festhalten der Ergebnisse der Validierung mittels JHOVE sowie des MD5-Wertes der Datei ist umfangreich, jedoch gewährt sie keinen Einblick in die Historie der Identifizierung und Validierung der Datei. Es ist nicht erkennbar, ob es sich hierbei um die Ergebnisse im Rahmen des Ingests oder eines späteren Durchlaufs handelt. Zwar werden das Erstelldatum und das Datum der letzten Veränderung des Primärdatenobjektes in den Metadaten festgehalten, doch könnten zwischen der Erstellung und letzten Änderungen dutzende Re-Identifizierungen und -Validierungen liegen, allein aus den Metadaten wird dies nicht deutlich. Jedoch wären solche Informationen beispielsweise in Hinblick auf die zwischenzeitlich verwendete DROID Signature File oder Nutzung eines aktualisierten oder anderen Tools von Interesse, da eine vorherige Entscheidung für oder gegen eine Bestandserhaltungsmaßnahme in hohem Maße von deren Ergebnissen abhing. Abhilfe könnte ein Protokoll schaffen, das jedem Primärdatenobjekt beigelegt wird und mindestens die AID sowie den Namen des Objekts, die PUID, die Maßnahme, das Enddatum der Maßnahme, die Akteure (hierbei handelt es sich sowohl um den den Prozess begleitenden Archivar als auch um die verwendeten Tools¹⁰⁴), die Ergebnisse und den Verweis mittels AID auch auf die Repräsentation umfasst.¹⁰⁵

¹⁰³ Vgl. Gespräch mit Julia Bettinger, David Radlinger, Kai Naumann (alle LA BW), 11.02.2020; DIMAG-Wiki [intern], <https://dimag-wiki.la-bw.de/xwiki/bin/view/Spezifikation/Protokollierung> (20.03.2020).

¹⁰⁴ Vgl. auch die Entität „Agent“ in PREMIS Editorial Committee, Data Dictionary, S. 16.

¹⁰⁵ Je nach Definition der Maßnahme kann das Protokoll umfangreicher ausfallen. Definiert man die Migration als eine Maßnahme, sind in ihr die Formaterkennung, -konvertierung und -validierung und ihre Ergebnisse inbegriffen, die jedoch auch alle als einzelne Maßnahmen verstanden werden können. Problematisch könnten sich in diesem Fall Containerformate wie Shapefiles herausstellen, die

Denkbar, sofern technisch umsetzbar, wäre eine automatisierte Teilbefüllung der Felder anhand der in PAR hinterlegten Daten zur verwendeten Preservation Action.¹⁰⁶

Bei der Konzipierung des BEM hat man sich gegen ein Protokoll für jedes Primärdatenobjekt entschieden. Stattdessen wird bei einer Migration ein Migrationsbericht innerhalb der neuen Repräsentation neben den Primärdatenobjekten abgelegt, der für alle Primärdatenobjekte einer Repräsentation gilt.¹⁰⁷ Er umfasst unter anderem die AID, den Namen, die Checksumme, die PUID und das Validierungsergebnis des Ausgangsobjekts sowie alle diese Angaben für das durch die Migration neu entstandene Primärdatenobjekt. Dadurch lassen sich die vorherige Stufe und der Migrationsprozess einer Repräsentation nachvollziehen. Es ergibt sich somit eine Kette aus Migrationsprotokollen in den einzelnen Repräsentationen. Dies kann in der Zukunft jedoch zu einem Problem führen: Entscheidet man sich bei Repräsentation 20 dafür, die Repräsentationen 3 bis 18 zu löschen, so wird auch die Protokollkette unterbrochen, da Repräsentation 20 lediglich die Informationen zur Migration aus Repräsentation 19 enthält. Ein Lösungsansatz wäre, beim Löschen der Repräsentationen die Migrationsberichte zu erhalten und nur die Primärdatenobjekte zu löschen. Ein weiteres Problem bleibt dabei jedoch bestehen: die Unübersichtlichkeit. Bei der alleinigen Ablage der Migrationsberichte in den Repräsentationen mutet man den Nutzern ein endloses Durchklicken durch Bericht für Bericht über die einzelnen Repräsentationen hinweg zu, ohne ein Gesamtbild der Maßnahmen zu präsentieren. Ein zusätzliches Gesamtprotokoll auf Ebene des Informationsobjektes würde hier Abhilfe verschaffen.¹⁰⁸ Dieses könnte summarisch die einzelnen Bestandserhaltungsschritte (nicht nur Migrationen) in der Geschichte dieses Informationsobjektes mit Datum, Charakterisierung der Maßnahme und den AIDs der betroffenen Primärdatenobjekte festhalten.¹⁰⁹ Den meisten Nutzern wird in der Rückschau eine Übersicht über die Maßnahmen auf Ebene des

letztendlich nicht nur aus einer, sondern aus mehreren Dateien bestehen. Da einzelne Protokolle für jede Datei unübersichtlich zu werden drohen, ist bei Containern die Dokumentationen der Maßnahmen in einem einzigen Protokoll mit Abschnitten zu jeder enthaltenen Datei sinnvoller.

¹⁰⁶ Infrage kämen hier der Name der Maßnahme, der damit einem kontrollierten Vokabular entspringen würde, sowie die verwendeten Tools. Es ist zu bedenken, dass innerhalb einer Maßnahme mehrere Tools verwendet und mehrere Ergebnisse festgehalten werden können. Bei einer Migration muss beispielsweise sowohl das Ausgangsformat mit einem Tool in das Zielformat überführt, als auch das Zielformat mit einem Tool validiert werden. Ein solches hier beschriebenes Protokoll wäre zum Festhalten aller möglichen Preservation Actions geeignet und nicht nur von Migrationen.

¹⁰⁷ Vgl. Gespräch mit Michael Tobegen, Staatsarchiv Hamburg, 27.02.2020; BEM-Benutzerhandbuch [intern], S. 2.

¹⁰⁸ Zuvor wäre zu prüfen, welche Belastung bei der Verarbeitung die große Zahl an Protokollen für DIMAG bedeuten würde und wieviel Speicherplatz benötigt würde.

¹⁰⁹ Durch die Angabe der AID der Repräsentation kann wiederum auf das ausführliche Protokoll verwiesen, am besten direkt verlinkt werden.

Informationsobjektes ausreichen, nur wenige werden sich für die Details der Veränderungen von Repräsentation zu Repräsentation interessieren.

Es erscheint darüber hinaus sinnvoll, eine weitere übergreifende Dokumentation auf Ebene des Gesamtarchivs vorzunehmen. Hier könnten Entscheidungen und deren Begründungen für Bestandserhaltungsmaßnahmen, die das ganze digitale Magazin betreffen und über alle Bestände und Informationsobjekte hinweg durchgeführt werden, festgehalten werden. So könnte beispielsweise die Grundsatzentscheidung protokolliert werden, dass alle PDF/A-Dateien in das (fiktive) PDF/Z-Format migriert werden sollen. Hierfür kann eine inhaltliche Begründung hinterlegt werden, denkbar wäre auch eine Beschreibung der Business Rule. Jeder Eintrag sollte mit einem Identifier gekennzeichnet werden. In den Protokollen auf Informationsobjekts-, Repräsentations- und möglicherweise auch Primärdatenobjektsebene sollte jeweils auf die zutreffende allgemeingültige Bestandserhaltungsentscheidung mittels des Identifier verwiesen werden. Dadurch sollte es möglich sein, sich alle von der Erhaltungsentscheidung betroffenen Objekte auflisten zu lassen.

IX. Eine Ampel für bedrohte Formate

Die Kunst der Bestandserhaltung besteht darin, den richtigen Zeitpunkt zur Einleitung der notwendigen Schritte nicht zu verpassen. Jedoch fehlt es an den entsprechenden Warnsystemen. Die Digital Preservation Coalition (DPC)¹¹⁰ hat mit der „Bit List“ of Digitally Endangered Species¹¹¹ in Zusammenarbeit mit der Bestandserhaltungcommunity den Versuch gestartet, die Bedrohungslage unterschiedlicher digitaler Unterlagenarten zu bewerten und in Anlehnung an die Rote Liste der gefährdeten Arten der Weltnaturschutzunion IUCN in sechs Kategorien einzuteilen: geringes Risiko (lower risk), verletzlich (vulnerable), bedroht (endangered), stark bedroht (critically endangered), praktisch ausgestorben (practically extinct) und Materialien, die Anlass zur Sorge geben (of concern), aber von der BitList Jury noch nicht bewertet wurden.

¹¹⁰ Vgl. <https://www.dpconline.org/> (24.02.2020). Zur Entstehung und Struktur der Organisation, die anfangs vor allem durch britische Bibliotheken, Archive und Wissenschaftseinrichtungen, mittlerweile durch Mitglieder weltweit getragen wird, vgl. Harvey/Weatherburn, Preserving Digital Materials, S. 179f.

¹¹¹ Vgl. <https://www.dpconline.org/our-work/bit-list> (24.02.2020); Kilbride, William, The Bit List of Digitally Endangered Species, in: Naumann, Kai/Dahlmanns, Tim (Hrsg.), Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen. Tagungsdokumentation. Workshop des KLA-Ausschusses Digitale Archive, Bundesarchiv Koblenz, 7./8. November 2018, <https://www.bundesarchiv.de/DE/Content/Downloads/KLA/tagungsdokumentation-bestandserhaltung-dig-unterlagen.pdf?blob=publicationFile> (24.02.2020).

Der Ansatz der BitList sollte in einer vereinfachten Version auf Dateiformate übertragen und in ein Ampelsystem aus drei Kategorien bzw. Farben überführt werden. Grün könnte für ein geringes Risiko, Gelb für bedroht und Rot für praktisch ausgestorben stehen. Ein intensiver Austausch über die Definition der Kategorien und der Ordnungskriterien ist notwendig und wird vermutlich kontrovers ausfallen. Dabei ist zu diskutieren, ob man allein die Gefahr einer Obsoleszenz zum Maßstab nehmen möchte oder ob weitere Anforderungen an die Archivierungstauglichkeit einbezogen werden sollten.¹¹² In einem zweiten Schritt sollte eine Datenbank aufgebaut werden, in der Formate den drei Kategorien zugeordnet werden. Es ist zu erwarten, dass einige Formate der Farbe Grün, die große Masse der Formate Gelb und (wenn überhaupt) nur wenige Rot zugeordnet werden. Zur Kennzeichnung der Formate sollte auf die PUID von PRONOM zurückgegriffen werden. In DIMAG könnte in den Metadaten eines jeden Primärdatenobjektes die Ampelphase des Formates angezeigt werden. Dies ermöglicht einen schnellen Überblick über die Gefährdungslage eines Primärdatenobjektes. Zwar könnte die Ampelkennzeichnung als ein weiteres Metadatum in einer Fülle von Metadaten untergehen, doch zum einen unterscheidet sie sich schon durch die optische Darstellung von den textlichen Metadaten und zum anderen kann man durch die Anknüpfung an die Ampelphasen auf verfestigte Sehgewohnheiten aufbauen. Vor allem bietet dieses Verfahren die Chance, in der täglichen Arbeit den Blick für gefährdete Formate und somit auch gefährdete Informationen zu schärfen – auch bei jenen Mitarbeitern, die nicht primär mit der Materie der digitalen Bestandserhaltung in ihren Aufgaben befasst sind. So lange automatisierte Bestandserhaltungsverfahren noch nicht flächendeckend umgesetzt sind, ist der Erfolg der Bestandserhaltung in hohem Maße von der Aufmerksamkeit, den Kenntnissen und der Risikoeinschätzung der zuständigen Archivare abhängig, welche durch ein Ampelsystem stärker sensibilisiert werden könnten.¹¹³ Natürlich ist es denkbar, an die Ampelphasen weitere Abfragen und statistische Auswertungen zu knüpfen.

¹¹² Die Bewertung kann hier durchaus unterschiedlich ausfallen: Ein Format, das weiterhin von ausreichend Software zur Darstellung unterstützt wird, könnte Grün zugeordnet werden. Sofern man aber auch Kategorien wie den Verbreitungsgrad oder das Vorhandensein von nicht-proprietärer Software einbezieht, könnte es auch anders eingefärbt werden.

¹¹³ Auch das BEM ist in der ersten Version nur teilautomatisiert und auf das manuelle Auslösen aller notwendigen Schritte durch den Nutzer angewiesen. Demnach muss der Nutzer dem BEM die zu migrierenden AIDs mitteilen, es erfolgt kein automatischer Scan des Magazins. Die Automatisierung ist erst in einer zweiten Version angedacht, die von den praktischen Erfahrungen mit der ersten Version profitieren soll, vgl. Gespräch mit Michael Tobegen, Staatsarchiv Hamburg, 27.02.2020.

Die entscheidende Frage lautet jedoch, wer eine solche Liste ‚Digitally Endangered Formats‘ anlegen und verwalten sollte. Ideal wäre eine internationale Lösung, beispielsweise angesiedelt bei den britischen National Archives und direkt verknüpft mit PRONOM. Ein ähnliches Unterfangen ist in PRONOM mit der Funktionalität ‚format risk‘ tatsächlich auch angelegt, aber nie umgesetzt worden. Die National Archives haben davon Abstand genommen, da man fürchtete, den Unmut der großen Software- und Technikunternehmen auf sich zu ziehen, falls man sich zum ‚Henker‘ eines ihrer Formate aufschwinge.¹¹⁴ Die Bereitschaft, die Rolle einer ‚Weltpolizei‘ für Formate einzunehmen, wird bei allen Institutionen wohl nur begrenzt vorhanden sein, da die Verabschiedung einer Risikoliste für Formate immer ein Politikum darstellt. Daher ist auch auf nationaler Ebene wie im nestor-Netzwerk mit ähnlichen Bedenken zu rechnen. Zumal eine Einigung bei der Bewertung einer Vielzahl von Formaten vermutlich nur schwer zu erzielen sein wird, wie schon die Divergenzen in den Whitelists zeigen.

Derzeit scheint allein eine Lösung auf regionaler bzw. lokaler Ebene innerhalb einer Archivverwaltung realistisch. Die technischen Voraussetzungen zur Implementierung eines Ampelsystems könnten zentral von einem Entwicklungspartner innerhalb des DIMAG-Verbundes umgesetzt werden, während jedes Landes-, Kommunalarchiv oder jeder weitere Anwendungspartner die Listen intern und individuell gestalten. Dies schließt den Austausch und die Vernetzung innerhalb des DIMAG-Verbundes keinesfalls aus, es wäre sogar höchst förderlich. Es ist aber anzunehmen, dass die meisten Involvierten von einer Veröffentlichung ihrer Listen zurückschrecken. Zum Aufbau der Datenbank könnte man sich zunächst auf gängige Formate beschränken, dann auf die restlichen Formate konzentrieren, die sich im digitalen Magazin zu

¹¹⁴ Die Funktion ist in den Hilfestellungen auf der PRONOM Website erläutert und mit Beispielen versehen worden, jedoch kann über die Suche ‚file format risk‘ keine Datensatz gefunden werden, der tatsächlich eine Eintragung bei ‚format risk‘ enthält – noch nicht einmal in dem Datensatz, der in dem Beispiel angebracht wird. Vgl. <http://www.nationalarchives.gov.uk/help/PRONOM/reportfileformat.htm> (17.03.2020); <http://www.nationalarchives.gov.uk/help/PRONOM/searchbyformat.htm> (17.03.2020). Laut David Clipsham (PRONOM, Gespräch vom 17.02.2020) wurde das ‚format risk‘ in PRONOM nie ausgefüllt. Auch die Funktion der ‚migration pathways‘ wurde mindestens in den letzten zehn Jahren nicht mehr gepflegt. Die fehlende (maschinenlesbare) Risikobewertung in PRONOM stellte Anfang der 2000er-Jahre auch den Automatic Obsolescence Notification Service (AONS), vor Schwierigkeiten, der von der University of Queensland, der Australian National University und der National Library of Australia entwickelt wurde. Dieser sollte auf Grundlage der Risikobewertungen externer Formatregister das digitale Magazin überwachen und bei entstehenden Obsoleszenzrisiken den Nutzer informieren. Vgl. Pearson, David, AONS II. Continuing the Trend Towards Preservation Software ‚Nirvana‘, in: Proceedings of iPRES 2007 4th International Conference on Preservation of Digital Objects (2007), <https://pdfs.semanticscholar.org/163d/d56e1b160cd1b20dd5267d689f71d063760c.pdf> (17.03.2020); Becker u.a., Preservation Watch.

diesem Zeitpunkt befinden. Diese können durch die Beschäftigung mit weiteren Formaten insbesondere im Prozess der Überlieferungsbildung ergänzt werden. Eine systematische Überarbeitung und Erweiterung sollte idealerweise jährlich erfolgen.

Es bleibt anzumerken, dass die Ampel keinen Automatismus und Aktionismus auslösen muss. Das Umspringen auf Rot lässt jedem Nutzer die Wahl, ob er tätig werden möchte oder nicht. Die Ampel kann aber leisten, dass sich dieser Nutzer überhaupt mit der Problematik beschäftigt.

X. Schlussbetrachtungen

Die Beobachtungsaufgaben sind in der Theorie grundlegend beschrieben, jedoch hat sich noch keine umfassende Praxis etabliert. Von einer systematischen, teilweise automatisierten Überwachung inklusive Warnsystemen ist man in einigen Bereichen der digitalen Bestandserhaltung weiter entfernt als in anderen. Es sind allerlei Hilfsmittel zur Hand: von Formatregistern, Toolregistern über neue Bestandserhaltungsmaßnahmenregister hin zu verschiedenen Tools in der Arbeit mit Formaten. Sie sortieren, stellen Informationen zur Verfügung und geben Orientierung, oftmals mangelt es aber an der Verzahnung; generell stellen sie kein Warnsystem dar. Die vorgeschlagene Ampel kann wiederum als Zwischenschritt verstanden werden: Auf der einen Seite wird eine Einschätzung des Obsoleszenzrisikos geboten, auf der anderen Seite muss die Schaltung der Ampelphasen weiterhin durch den Archivar per Hand vorgenommen werden. Des Weiteren wird kein Archiv mehr ohne einen Katalog langzeitarchivierungsfähiger Formate auskommen, da sie zu einer Komplexitätsreduktion beitragen, ohne die auf lange Sicht eine Bestandserhaltung nicht durchführbar scheint.

Der Blick auf die Formaterkennung und DIMAG-Statistik hat gezeigt, dass jede Bestandserhaltungsplanung nur so gut sein kann, wie die ihr zugrundeliegenden Daten aktuell sind. Sowohl veraltete, als auch irrelevante Identifizierungsergebnisse der nicht aktuellsten Repräsentation können zu falschen Rückschlüssen führen, weshalb Re-Identifizierungen und -validierungen anlassbezogen und periodisch durchgeführt werden sollten. Dieser Zweck erfordert differenziertere Abfragemöglichkeiten und eine transparente Dokumentation der Erhaltungsentscheidungen und -Maßnahmen in DIMAG.

Diese Arbeit setzte implizit die Migrationsstrategie als führende Erhaltungsstrategie voraus. Jedoch sollte in Anbetracht des zunehmenden Wildwuchses an Formaten die Emulation in den Überlegungen der Archive einen größeren Raum einnehmen, insbesondere bei eher exotischen Dateiformaten. Die Entscheidung für ein

Verfahren und den damit verbundenen Aufwand wird stark von der Designated Community und dem Archivalientyp abhängen. Es ist vorstellbar, dass weniger Energie in die Erhaltung eines bestimmten Archivalientyps in einem besonderen Format aufgewendet wird, da die entsprechende Designated Community sich als hochtechnisiert erwiesen hat, die sich selber Emulationslösungen erarbeitet. Die Ausrichtung der Entscheidungen an den Fähigkeiten, Ansprüchen und Nutzungen der Designated Community ist völlig angemessen. Bestandserhaltungsmaßnahmen ‚am Markt vorbei‘ könnten sich als aktive Unbrauchbarmachung der Informationen herausstellen. Dennoch besteht die Gefahr, dass sich die tatsächliche oder zu erwartende Nutzung eines Bestandes zum alleinigen Kriterium für eine Bestandserhaltungsmaßnahme entwickelt.¹¹⁵ Die Interessen und Bedeutung der unterschiedlichen Nutzergruppen könnten gegeneinander ausgespielt und der Informationswert, die Bedeutung des Archivals als solches vergessen werden. Die Notwendigkeit zur Priorisierung ist ein Eingeständnis an begrenzte Ressourcen und betrifft die analoge ebenso wie die digitale Bestandserhaltung. Die Entscheidung, *ob* etwas erhalten wird, sollte dabei keinesfalls primär von der Designated Community abhängig gemacht werden, *wie* erhalten wird, hingegen schon.

Es gilt, das richtige Maß aus Aktionismus und beobachtendem Warten¹¹⁶ zu finden. Die Formatobsoleszenz hat sich bisher als nicht so gravierend wie befürchtet herausgestellt, ignorieren darf man sie hingegen auch nicht. Policies, Abläufe, Tools, Anforderungen und Funktionalitäten sollten frühzeitig diskutiert, entwickelt, erprobt und implementiert werden. Insbesondere die Prozesse und Anwendungen zur Massmigration gilt es voranzutreiben – mit dem BEM ist ein Schritt getan. Die Entwicklungen können rasant sein, aber nicht so rasant, als dass keine Zeit für einen intensiven Austausch und eine Zusammenarbeit mit Experten, Kollegen und Kolleginnen in der internationalen Bestandserhaltungsgemeinschaft bestünde.

¹¹⁵ Mario Glauert hat dieses Argument insbesondere für den Bereich der Überlieferungsbildung und Erschließung entwickelt: „Sie [die Nutzung] könnte auch zum entscheidenden Prioritätskriterium, faktisch zu einer zweiten Bewertung für die künftige Auswahl von Archivalien und Beständen werden, wenn es um ihre dauerhafte Erhaltung und digitale Bereitstellung geht. Das Revolutionäre an dieser neuen Form der Überlieferungsbildung ist allerdings, dass sie nicht mehr archivar-, sondern nutzergesteuert ist.“, Glauert, Mario, Was ist ein erfolgreiches Archiv? Archivbenutzung im Zeitalter der digitalen Transformation, in: Brandenburgisches Landeshauptarchiv (Hrsg.), Arbeiten für das Gedächtnis des Landes. Übernehmen – erschließen – auswerten – bewahren – bereitstellen. Festgabe zum 70-jährigen Jubiläum des Brandenburgischen Landeshauptarchivs, Potsdam 2019, S. 10–17, hier S. 16.

¹¹⁶ Beim „Watchful waiting“ handelt es sich um eine Behandlungsstrategie insbesondere bei Prostatakrebs, bei der zunächst keine Therapiemaßnahmen durchgeführt werden, sondern der Krankheitsverlauf überwacht wird. S.a. Rosenthal, Format Obsolescence.

Zusammenfassung

Der Erhalt digitaler Informationen erfordert eine aufmerksame Beobachtung der Umwelt eines digitalen Magazins, sowohl in Bezug auf die Technologie als auch auf die Designated Community. Insbesondere die Obsoleszenz von Dateiformaten bedroht die Les- und Darstellbarkeit von Informationen. Formatregister, Kataloge archivierungsfähiger Dateiformate sowie Tools zur Formaterkennung und -validierung stellen wichtige Hilfsmittel dar, doch mangelt es sowohl an einer verstärkten Verzahnung und Automatisierung von Maßnahmen als auch an einem umfassenden Warnsystem. Die Erhaltungsplanung wird durch veraltete Formaterkennungsergebnisse und eingeschränkte Abfragemöglichkeiten innerhalb DIMAGs erschwert. Es ist deutlich geworden, dass Archive als Einzelkämpfer die umfangreichen Aufgaben der digitalen Bestandserhaltung nicht bewältigen können, weshalb mit Blick auf Initiativen wie PRO-NOM und den Preservation Action Registries die Notwendigkeit einer engeren nationalen wie internationalen Zusammenarbeit betont wurde.

Anhang

Anlage 1: Archivierungsfähige Dateiformate

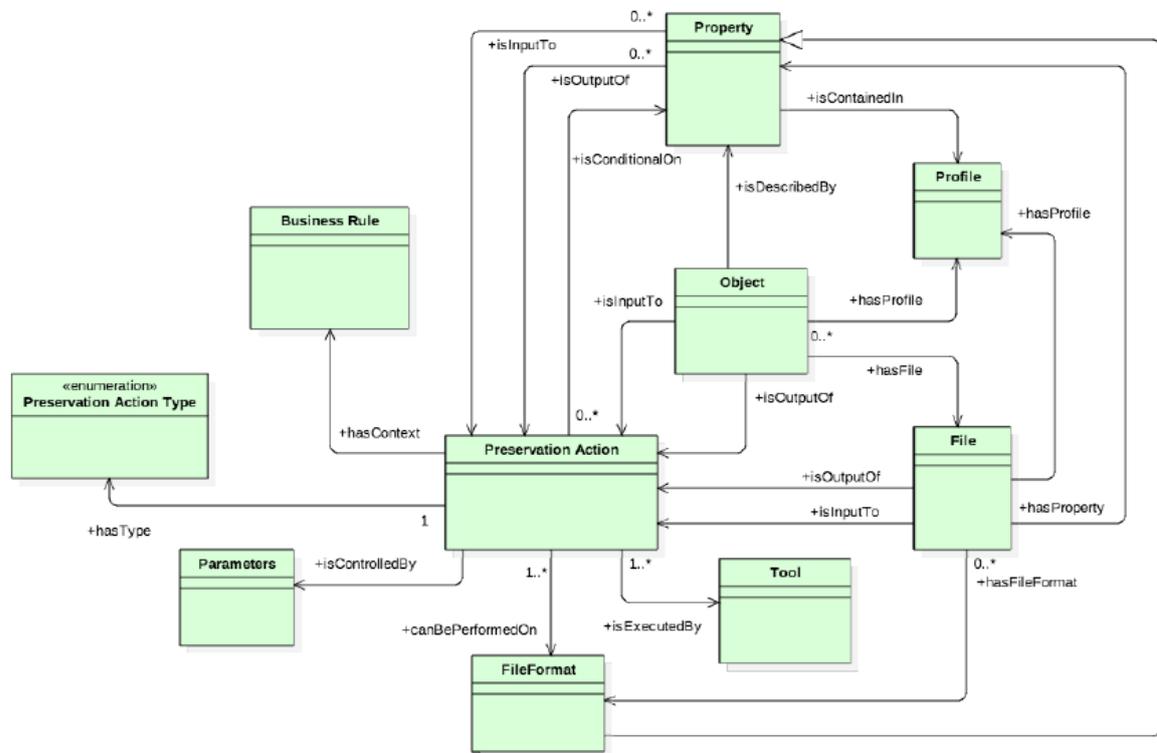
| | Text | Rastergrafiken | Audio | Video | Strukturierte Daten (Tabellenkalkulation) | Datenbanken | Webarchivierung / Hyper-text | GIS | CAD/CAM, Vektorgrafiken |
|-------------------------------------|--|---|--|---|--|--|---|--|---|
| KOST | <i>Bevorzugt:</i> PDF/A <i>Weitere:</i> PDF Plain text | <i>Bevorzugt:</i> TIFF JPEG2000 DNG PDF/A <i>Weitere:</i> PNG | <i>Bevorzugt:</i> WAVE <i>Weitere:</i> FLAC ALAC | <i>Noch keine abschließende Empfehlung, Zurückhaltung bei der Migration bevorzugt.</i> <i>Weitere:</i> Digital Video FFV1 MPEG-2 MPEG-4 MJPEG2000 | <i>Bevorzugt:</i> XLS ODF OOXML | <i>Bevorzugt:</i> SIARD CSV <i>Weitere:</i> SQLX | <i>Bevorzugt:</i> PDF/A <i>Weitere:</i> WARC | <i>Bevorzugt:</i> INTERLIS Shapefile <i>Weitere:</i> GeoTIFF Rasterdaten mit World-File | <i>Bevorzugt:</i> DWG <i>Weitere:</i> IFC Vektorgrafik in PDF |
| Schweizerisches Bundesarchiv | Plain text PDF/A XML/XSD | TIFF JPEG2000 | WAVE | MPEG-4 FFV1 | CSV | SIARD | PDF/A SIARD XML | TIFF + EWF.XML INTERLIS | |
| ETH Zürich | <i>Bevorzugt:</i> PDF/A XML XSD XSL XHTML Unformatierter Text <i>Weitere:</i> PDF RTF HTML | <i>Bevorzugt:</i> TIFF PNG JPEG2000 (verlustfrei) DNG <i>Weitere:</i> TIFF (komprimiert) GIF BMP JPEG/JFIF JPEG2000 (verlusthaft) | <i>Bevorzugt:</i> WAVE <i>Weitere:</i> MP4 MP3 | <i>Bevorzugt:</i> FFV1 <i>Weitere:</i> MPEG-2 MPEG-4 MP4 MOV AVI Motion JPEG2000 | <i>Bevorzugt:</i> CSV <i>Weitere:</i> XLSX ODM ODT ODG ODC ODF | <i>Bevorzugt:</i> CSV | | | <i>Bevorzugt:</i> DWG DXF X3D X3DV X3DB SVG |

| | Text | Rastergrafiken | Audio | Video | Strukturierte Daten (Tabellenkalkulation) | Datenbanken | Webarchivierung / Hypertext | GIS | CAD/CAM, Vektorgrafiken |
|--------------------------------|--|---|---|---|---|---|---|---|---|
| LOC | <p><i>Bevorzugt:</i> EPUB3 PDF/UA PDF/A PDF UTF-8 UTF-16 (with BOM) US-ASCII</p> <p><i>Weitere:</i> XHTML HTML DOCX/OOXML 2012 ODF OOXML SGML PDF RTF Plain text</p> | <p><i>Bevorzugt:</i> TIFF JPEG2000 PNG JPEG/JFIF DNG BMP GIF</p> <p><i>Weitere:</i> Photoshop (*.psd, *.psb) JPEG2000 Part 2 (*.jpf, *.jpx) Proprietary Camera Raw formats (*.nef, *.crw)</p> | <p><i>Bevorzugt:</i> PCM WAVE Broadcast WAVE DSD PCM DFF DSF</p> <p><i>Weitere:</i> Highest resolution compressed version in a major standard compression scheme Lossless compression scheme rather than lossy compression scheme</p> | <p><i>Bevorzugt:</i> DCP IMF Pro-Res MPEG-2 XDCAM</p> <p><i>Weitere:</i> MPEG-4</p> | <p><i>Bevorzugt:</i> XLS</p> | <p><i>Bevorzugt:</i> CSV SQLite</p> | <p><i>Bevorzugt:</i> WARC</p> <p><i>Weitere:</i> ARC_IA</p> | <p><i>Bevorzugt:</i> GeoTIFF GeoPDF GeoJPEG2000 Shapefile</p> | <p><i>Bevorzugt:</i> SVG DXF Shapefile</p> <p><i>Weitere:</i> CGM, WebCGM IFC STEP PDF/UA PDF/A PDF PDF/X EPS</p> |
| Harvard Digital Library | <p><i>Bevorzugt:</i> PDF PDF/A PDF/X DOC DOCX RTF WPD</p> | <p><i>Bevorzugt:</i> TIFF JPEG2000 JPEG JFIF GIF</p> <p><i>Weitere:</i> JPEG TIFF mit Komprimierung</p> | <p><i>Bevorzugt:</i> WAVE MP4</p> <p><i>Weitere:</i> AIFF</p> | <p><i>Bevorzugt:</i> JPEG 2000 DV MPEG-2</p> <p><i>Weitere:</i> Avid DNxHD Apple ProRes</p> | | | | | <p><i>Bevorzugt:</i> PDF DWG AutoCAD DXF X3D</p> |

| | Text | Rastergrafiken | Audio | Video | Strukturierte Daten (Tabellenkalkulation) | Datenbanken | Webarchivierung / Hyper-text | GIS | CAD/CAM, Vektorgrafiken |
|------------------------------------|--|---|--|--|---|-------------|---|---|--|
| Library and Archives Canada | <p><i>Bevorzugt:</i> ASCII Text EPUB3 ODF PDF/A Unicode Text</p> <p><i>Weitere:</i> EPUB2.0.1 DOC DOCX PDF</p> | <p><i>Bevorzugt:</i> TIFF JPEG2000 PNG</p> <p><i>Weitere:</i> JPEG/JFIF DNG GIF DICOM</p> | <p><i>Bevorzugt:</i> BWF</p> <p><i>Weitere:</i> AIFF MP3 AAC WAVE</p> | <p><i>Bevorzugt:</i> DCDM DPX AVI (unkomprimiert) MXF MOV (unkomprimiert)</p> <p><i>Weitere:</i> DCP AVI MOV MPEG2 MPEG4 WMV</p> | <p><i>Bevorzugt:</i> ASCII Text CSV</p> <p><i>Weitere:</i> EBCDIC DBF OOXML XLS ODS</p> | | | <p><i>Bevorzugt:</i> BIL BSQ DEM Shapefile GeoTIFF GML KML</p> <p><i>Weitere:</i> CCOGIF DIG-3 GeoPDF</p> | <p><i>Bevorzugt:</i> DXF</p> <p><i>Weitere:</i> PDF/E STEP</p> |
| National Archives (USA) | <p><i>Bevorzugt:</i> ASCII Text ODF PDF/A Unicode Text</p> <p><i>Weitere:</i> DOC DOCX PDF</p> | <p><i>Bevorzugt:</i> TIFF</p> <p><i>Weitere:</i> DNG JPEG/JFIF JPEG2000 PNG</p> | <p><i>Bevorzugt:</i> BWF FLAC</p> <p><i>Weitere:</i> AIF MPEG Audio Layer III (MP3) WAVE</p> | <p><i>Bevorzugt:</i> DPX</p> <p><i>Weitere:</i> AVI (unkomprimiert) MOV (unkomprimiert) MPEG-2 MPEG-4 MXF WMV</p> | <p><i>Bevorzugt:</i> ASCII Text CSV JSON ODS XML</p> <p><i>Weitere:</i> OOXML XLS</p> | | <p><i>Weitere:</i> ARC WARC</p> | <p><i>Bevorzugt:</i> GeoTIFF GML</p> <p><i>Weitere:</i> GeoPDF Shapefile</p> | <p><i>Bevorzugt:</i> STEP X3D</p> <p><i>Weitere:</i> PDF/E PRC U3D</p> |
| National Archives (UK) | <p>DOC DOCX DOT ODT PDF PDF/A PDF/E PDF/X RTF SXW WP WPD XML XSD</p> | <p>BMP CDR JPEG JPEG2000 ODG PNG SXD TIFF</p> | <p>MID MPEG 1/2 Audio Layer 3 MPEG-4 OGG WAVE WMA</p> | <p>MJPEG2000 MOV MPEG-1 MPEG-2 MPEG-4 MXF SWF WEBM WMV</p> | <p>CSV ODS SXC WKS WKS XLS XLSB XLSX</p> | | | | |

| | Text | Rastergrafiken | Audio | Video | Strukturierte Daten (Tabellenkalkulation) | Datenbanken | Webarchivierung / Hypertext | GIS | CAD/CAM, Vektorgrafiken |
|----------------------------------|---|--|--|---|--|--|-----------------------------|------------------------|-------------------------|
| Landesarchiv Nordrhein-Westfalen | PDF/A | PDF/A | <i>Keine allgemeine Empfehlung.</i> | <i>Keine allgemeine Empfehlung.</i> | <i>Kein Langzeitarchivierungsformat vorhanden, dass alle signifikanten Eigenschaft abdeckt.</i> PDF/A + Ursprungsformat | | | | |
| Landesarchiv Hessen | PDF/A Plain Text | TIFF | WAVE | MPEG-2 | CSV XML | CSV XML | | | SVG |
| Landesarchiv Rheinland-Pfalz | DOC DOCX PDF PDF/A RTF XML | JPEG2000 | WAVE | MPEG-4 | CSV XLSX | CSV | | | |
| Landesarchiv Thüringen | PDF/A XML | <i>Bevorzugt:</i> TIFF <i>Weitere:</i> BMP JPEG JPEG2000 PNG | <i>Bevorzugt:</i> FLAC WAVE <i>Weitere:</i> MP3 OGG | <i>Bevorzugt:</i> MJPEG2000 <i>Weitere:</i> AVI FLV MJPEG WMV | <i>Bevorzugt:</i> CSV XML <i>Weitere:</i> ODS XLS XLSX | <i>Bevorzugt:</i> CSV Plain Text SIARD XML <i>Weitere:</i> SQL | | <i>Nach Absprache.</i> | |

Anlage 2: PAR Conceptual Model



PAR Conceptual Model

Abb. in: Addis, Matthew/Simpson, Justin/Tilbury, Jonathan/O'Sullivan, Jack/Stokes, Paul, Digital Preservation Interoperability through Preservation Actions Registries (2018).
<https://doi.org/10.6084/m9.figshare.6628418.v1> (27.01.2020).

Anlage 3: PAR Business Rule Eintrag

archivematica Transfer Backlog Appraisal Ingest Archival storage Preservation planning Access Administration test

PAR Preservation Actions / List

FPR PAR

Business Rules | Preservation Actions

description EBU Core XML is the required format for recording technical characteristics of Quicktime Movie files.

id e42aa5b2-5652-4a01-a037-7d0ae817a73d

notes Modern Institution has decided to adopt EBUCore as the primary metadata schema for describing technical characteristics of Quicktime Video Files. Only EBUCore metadata is allowed. See http://modern-institution.edu/preservation_policies

preservationActionTypes metadata extraction (<http://id.loc.gov/vocabulary/preservation/eventType/mee>)

preservationActions MedialInfo3 (1caa0cde-e345-44ac-8d83-51afaa7427b6)

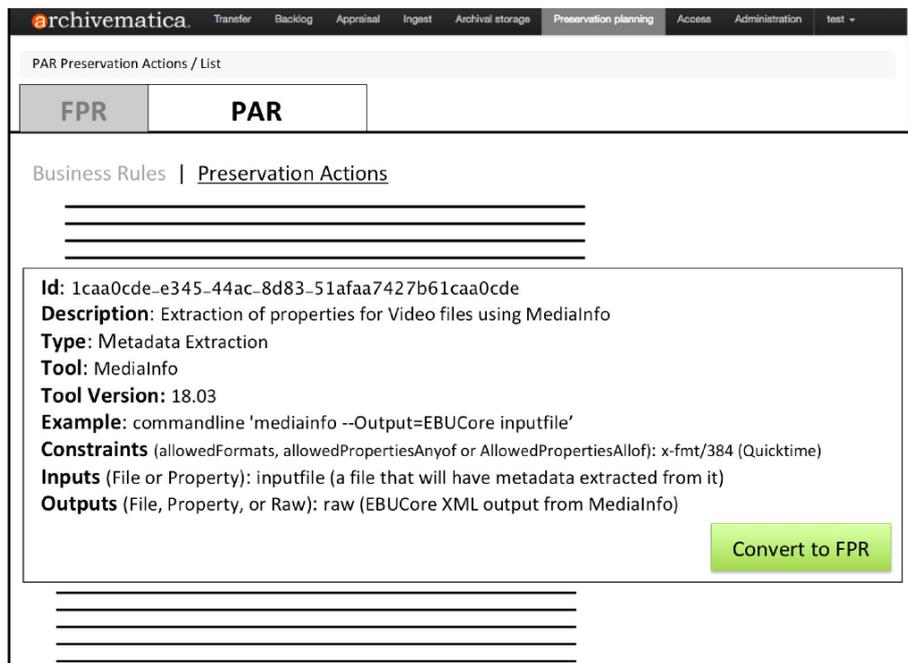
priority 1

rawOutputsRetrieved EBUCore XML output from MedialInfo

Convert to FPR

Abb. in: Präsentation der OPF: „Preservation Action Registries Workshop“, iPres Event Workshop, September 2019, <http://parcore.org/presentations/PAR%20Presentation%20iPres%202019.pdf> (12.03.2020), S. 26.

Anlage 4: PAR Preservation Action Eintrag



The screenshot displays the Archivematica web interface. At the top, a navigation bar includes the Archivematica logo and menu items: Transfer, Backlog, Appraisal, Ingest, Archival storage, Preservation planning (active), Access, Administration, and test. Below the navigation bar, the page title is "PAR Preservation Actions / List". There are two tabs: "FPR" and "PAR", with "PAR" selected. The main content area shows "Business Rules | Preservation Actions". A list of actions is shown, with the first one expanded to show details:

- Id:** 1caa0cde-e345-44ac-8d83-51afaa7427b61caa0cde
- Description:** Extraction of properties for Video files using MediaInfo
- Type:** Metadata Extraction
- Tool:** MediaInfo
- Tool Version:** 18.03
- Example:** commandline 'mediainfo --Output=EBUCore inputfile'
- Constraints** (allowedFormats, allowedPropertiesAnyof or AllowedPropertiesAllof): x-fmt/384 (Quicktime)
- Inputs** (File or Property): inputfile (a file that will have metadata extracted from it)
- Outputs** (File, Property, or Raw): raw (EBUCore XML output from MediaInfo)

A green button labeled "Convert to FPR" is located at the bottom right of the expanded entry. Below the entry, there are several horizontal lines representing other entries in the list.

Abb. in: Präsentation der OPF: „Preservation Action Registries Workshop“, iPres Event Workshop, September 2019, (12.03.2020), S. 27.

Abkürzungsverzeichnis

| | |
|--------|--|
| AKEA | Arbeitskreis „Elektronische Archivierung“ der Vereinigung deutscher Wirtschaftsarchive |
| AONS | Automatic Obsolescence Notification Service |
| BEM | DIMAG-Bestandserhaltungsmodul |
| CCSDS | Consultative Committee of Space Data Systems |
| COPTR | Community Owned digital Preservation Tool Registry |
| DAN | Digitale Archivierung Nord |
| DIMAG | Digitales Magazin des Landesarchiv Baden-Württembergs |
| DPC | Digital Preservation Coalition |
| DROID | Digital Record Object Identification |
| FIDO | Format Identification for Digital Objects |
| FPR | Format Policy Register von Archivemata |
| GDFR | Global Digital Format Registry |
| IANA | Internet Assigned Numbers Authority |
| IUCN | International Union for Conservation of Nature |
| JHOVE | JSTOR/Harvard Object Validation Environment |
| JISC | Joint Information Systems Committee |
| KaD | Katalog archivischer Dateiformate der KOST |
| KOST | Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen |
| LA BW | Landesarchiv Baden-Württemberg |
| LDR | Linked Data Registry von Preservica |
| LOC | Library of Congress |
| OAIS | Reference Model for an Open Archival Information System |
| OPF | Open Preservation Foundation |
| PAR | Preservation Action Registries |
| PREMIS | Preservation Metadata: Implementation Strategies |
| PUID | PRONOM Persistent Unique Identifier |
| RDSS | Research Data Shared Service |
| UDFR | Unified Digital Format Registry |

Literatur- und Quellenverzeichnis

- Addis, Matthew/Simpson, Justin/Tilbury, Jonathan/O’Sullivan, Jack/Stokes, Paul, Digital Preservation Interoperability through Preservation Actions Registries (2018), <https://doi.org/10.6084/m9.figshare.6628418.v1> (27.01.2020).
- Apache Tika, <http://tika.apache.org/> (06.03.2020).
- Archivemata, Preservation Planning, <https://www.archivemata.org/en/docs/archivemata-1.10/user-manual/preservation/preservation-planning/> (11.03.2020).
- Arms, Caroline/Fleischhauer, Carl, Digital Formats. Factors for Sustainability, Functionality and Quality (2005), https://memory.loc.gov/ammem/techdocs/digform/Formats_IST05_paper.pdf (24.02.2020).
- Aschenbrenner, Andreas/Wollschläger, Thomas, File Format Registries, in: Neuroth, Heike/Oßwald, Achim/Scheffel, Regine/Huth, Karsten (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010), <http://nbn-resolving.de/urn:nbn:de:0008-2010071949> (28.01.2020), Kap. 7:19–Kap. 7:22.
- Bachmann, Steffen/Ernst, Katharina, Formaterkennung. Ziele, Herausforderungen, Lösungsansätze, in: Manke, Matthias (Hrsg.), Auf dem Weg zum digitalen Archiv. Stand und Perspektiven von Projekten zur Archivierung digitaler Unterlagen. 15. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 2. und 3. März 2011 in Schwerin, Schwerin 2012, S. 69–73.
- Becker, Christoph/Duretec, Kresimir/Petrov, Petar/Faria, Luis/Ferreira, Miguel/Ramalho, Jose Carlos, Preservation Watch. What to Monitor and How, in: Proceedings of iPRES 2012 9th International Conference on Preservation of Digital Objects (2012), http://publik.tuwien.ac.at/files/PubDat_213865.pdf (14.02.2020).
- BEM-Benutzerhandbuch, Version vom 23.05.2019 [DIMAG-Wiki, intern], https://dimag-wiki.la-bw.de/xwiki/bin/download/DIMAG-Wiki+Startbereich/Bestandserhaltungsmodul/BEM_Benutzerhandbuch_2019-05-23.pdf (22.03.2020).
- Blekinge, Asgar, Identification Tools, an Evaluation. The Scape Characterisation Tool Testing Suite (2012), <https://openpreservation.org/blog/2012/02/23/identification-tools-evaluation/> (14.02.2020).
- Braud, Maïté/Sinclair, Pauline/Sharpe, Robert, Linked Data Registry. A New Approach to Technical Registries; Archives and Cultural Industries, 11th–15th October 2014, <https://pdfs.semanticscholar.org/98bd/5a5ac26c40438b8af978021bdca8123f1988.pdf> (11.03.2020).
- Brown, Adrian, Practical Digital Preservation. A How-To Guide for Organizations of Any Size, London 2013.
- Caplan, Priscilla, PREMIS verstehen (2017), https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/berichte/pREMISverstehen2017.pdf?__blob=publicationFile&v=1 (20.01.2020).
- Community Owned digital Preservation Tool Registry (COPTR), https://coptr.digipres.org/Main_Page (06.03.2020).

- Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS). Magenta Book (CCSDS 650.0-M-2) (2012), <https://public.ccsds.org/Pubs/650x0m2.pdf> (15.01.2020).
- Corrado, Edward M./Moulaison Sandy, Heather, Digital Preservation for Libraries, Archives, and Museums, 2. Aufl., Lanham u.a. 2017.
- Digital Preservation Coalition (DPC), Digital Preservation Handbook, 2. Aufl. 2015, <https://www.dpconline.org/handbook> (20.03.2020).
- Digital Preservation Coalition, <https://www.dpconline.org/> (24.02.2020).
- DIMAG Wiki, Protokollierung [intern], <https://dimag-wiki.la-bw.de/xwiki/bin/view/Spezifikation/Protokollierung> (20.03.2020).
- ETH Zürich, Archivtaugliche Dateiformate, <https://documentation.library.ethz.ch/display/RC/Archivtaugliche+Dateiformate#space-menu-link-content> (28.02.2020).
- FITS, <https://projects.iq.harvard.edu/fits> (06.03.2020).
- Funk, Stefan E./Neubauer, Matthias, 7.4 Formatcharakterisierung, in: Neuroth, Heike/OBwald, Achim/Scheffel, Regine/Huth, Karsten (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010), <http://nbn-resolving.de/urn:nbn:de:0008-2010071949> (28.01.2020), Kap. 7:13–Kap. 7:15.
- Glauert, Mario, Was ist ein erfolgreiches Archiv? Archivbenutzung im Zeitalter der digitalen Transformation, in: Brandenburgisches Landeshauptarchiv (Hrsg.), Arbeiten für das Gedächtnis des Landes. Übernehmen – erschließen – auswerten – bewahren – bereitstellen. Festgabe zum 70-jährigen Jubiläum des Brandenburgischen Landeshauptarchivs, Potsdam 2019, S. 10–17.
- Griffin, Andrew, ‚Digital Dark Age‘ Could Leave Historians With No Records of the 21st Century, in: Independent 13.02.2015, <https://www.independent.co.uk/life-style/gadgets-and-tech/news/digital-dark-age-could-leave-historians-with-no-records-of-the-21st-century-10043516.html> (19.03.2020).
- Gutzmann, Ulrike/Kamp, Ulrich/Keitel, Christian/Scheidung, Antje, Praktische Lösungsansätze zur Archivierung digitaler Unterlagen. „Langzeitarchivierung“ und dauerhafte Sicherung der digitalen Überlieferung, in: Archiv und Wirtschaft. Zeitschrift für das Archivwesen der Wirtschaft 40 (2007), S. 20–27.
- Harvard University Library, Formats Supported by the DRS, <https://wiki.harvard.edu/confluence/display/digitalpreservation/Formats+Supported+by+the+DRS> (28.02.2020).
- Harvey, Ross/Weatherburn, Jaye, Preserving Digital Materials, 3. Aufl., Lanham/London 2018.
- Hessisches Landesarchiv, Archivfähige Dateiformate und ihre Spezifikationen (2016), <https://landesarchiv.hessen.de/sites/landesarchiv.hessen.de/files/content-downloads/Archivf%C3%A4hige+Dateiformate+%28Stand+August+2016%29.pdf> (02.03.2020).
- Internet Assigned Numbers Authority (IANA), Media Types, <https://www.iana.org/assignments/media-types/media-types.xhtml> (04.03.2020).

- JHOVE, <http://jhove.openpreservation.org/> (06.03.2020).
- Karaiskos, Christos, Understanding Apple's Binary Property List Format (27.02.2018), <https://medium.com/@karaiskc/understanding-apples-binary-property-list-format-281e6da00dbd> (10.03.2020).
- Keitel, Christian, Authentische Archive. Wunsch und Wirklichkeit, in: Herrmann, Tobias (Hrsg.), Verlässlich, richtig, echt – Demokratie braucht Archive! 88. Deutscher Archivtag in Rostock, Fulda 2019, S. 123–131.
- Keitel, Christian, Benutzerinteressen annehmen und signifikante Eigenschaften festlegen. Einige neue Aufgaben für Archivare, in: Schmitt, Heiner (Hrsg.), Archive im digitalen Zeitalter. Überlieferung, Erschließung, Präsentation. 79. Deutscher Archivtag in Regensburg, Neustadt a.d. Aisch 2010, S. 29–42.
- Keitel, Christian, Der nestor-Leitfaden zur Digitalen Bestandserhaltung und seine Folgen für die Archive, in: ders. (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus Digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“, Stuttgart 2013, S. 267–280.
- Keitel, Christian, Zwölf Wege ins Archiv. Umriss einer offenen und praktischen Archivwissenschaft, Stuttgart 2018.
- Kilbride, William, The Bit List of Digitally Endangered Species, in: Naumann, Kai/Dahlmanns, Tim (Hrsg.), Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen. Tagungsdokumentation. Workshop des KLA-Ausschusses Digitale Archive, Bundesarchiv Koblenz, 7./8. November 2018, https://www.bundesarchiv.de/DE/Content/Downloads/KLA/tagungsdokumentation-bestandserhaltung-dig-unterlagen.pdf?__blob=publicationFile (24.02.2020).
- König, Mareike, Was sich Historiker*innen von Archiven wünschen: eine Umfrage (2019), <https://dhdhi.hypotheses.org/6107> (05.11.2019).
- Kortyla, Stephanie, Digitale Archivierung. (Zu-)Fälle bei der Formatidentifizierung am Beispiel von PRONOM und DROID, in: Sächsisches Archivblatt 2017, H. 1, S. 20f.
- Kortyla, Stephanie/Treu, Christian, Nutzen und Grenzen der Formatidentifizierung, in: Informationswissenschaft: Theorie, Methode und Praxis 5 (2018), S. 83–95.
- KOST, Katalog archivischer Dateiformate, https://kost-ceco.ch/cms/kad_main_de.html (20.03.2020).
- Landesarchiv Baden-Württemberg, Archivierung digitaler Unterlagen: Normen, Vorschriften, Richtlinien, <https://www.landesarchiv-bw.de/web/46913> (02.03.2020).
- Landesarchiv Nordrhein-Westfalen, Leitfaden für langzeitstabile Datenformate in der elektronischen Aktenführung. Version 1.1 (2015), https://www.archive.nrw.de/lav/abteilungen/fachbereich_grundsaeetze/BilderKartenLogosDateien/Behoerdeninformation/Bausteine_III_Langzeitstabile_Datenformate.pdf (01.03.2020).
- Landesarchiv Thüringen, Empfehlungen zur Anbietung und Übergabe elektronischer Unterlagen an die Thüringischen Staatsarchive. Version 3.3 (2013),

- https://www.thueringen.de/mam/th1/staatsarchive/behoerden/empfehlung_uebergabe_elektronischen_archivguts_v3.3.pdf (02.03.2020).
- Landesarchivverwaltung Rheinland-Pfalz, Liste der archivfähigen Dateiformate, https://www.landeshauptarchiv.de/fileadmin/user_upload/Gemeinsame_Dateien/Liste_der_archivf%C3%A4higen_Dateiformate.pdf (02.03.2020).
- Let's Solve the File Format Problem!, <http://fileformats.archiveteam.org/> (24.02.2020).
- Library and Archives Canada, Guidelines on File Formats for Transferring Information Resources of Enduring Value, <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx> (28.02.2020).
- Library of Congress, Preservation Vocab „Event Type“, <http://id.loc.gov/vocabulary/preservation/eventType.html> (11.03.2020).
- Library of Congress, Recommended Formats Statement 2019–2020, <https://www.loc.gov/preservation/resources/rfs/TOC.html> (28.02.2020).
- Library of Congress, Sustainability of Digital Formats, <https://www.loc.gov/preservation/digital/formats/index.html> (20.03.2020).
- Lindlar, Michelle/Tunnat, Yvonne, Time-Travel with PRONOM. The 4th Dimension of DROID, in: Proceedings of iPRES 2018 15th International Conference on Digital Preservation (2018), https://zenodo.org/record/3517767/files/iPres2018_paper_85_final.pdf?download=1 (27.01.2020).
- List of file formats, in: Wikipedia, https://en.wikipedia.org/wiki/List_of_file_formats (03.03.2020).
- Media type, in: Wikipedia, https://en.wikipedia.org/wiki/Media_type (04.03.2020).
- MIT Libraries, „Chamber of Horrors“, <https://dpworkshop.org/dpm-eng/oldmedia/chamber.html> (03.03.2020).
- Museum of Obsolete Media, <https://obsoletemedia.org/> (03.03.2020).
- National Archives, Appendix A: Tables of File Formats, <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html> (28.02.2020).
- Naumann, Kai, Zwölf Jahre Lernen aus der Praxis. Überlieferungsbildung aus genuin digitalen Unterlagen beim Landesarchiv Baden-Württemberg, in: *Scrinium* 69 (2015), S. 115–136.
- Nestor Wiki, <https://wiki.dnb.de/display/NESTOR/Startseite> (20.03.2020).
- nestor, File Test Suite, in: Github, <https://github.com/nestorFormatGroup/FileTestSuite> (06.03.2020).
- nestor, Referenzmodell für ein Offenes Archiv-Informations-System. Deutsche Übersetzung 2.0 (2013), <https://nbn-resolving.org/urn/resolver.pl?urn=urn:nbn:de:0008-2013082706> (15.01.2020).
- nestor-Arbeitsgruppe Digitale Bestandserhaltung, Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 2.0 (2012), <http://nbn-resolving.de/urn:nbn:de:0008-2012092400> (22.01.2020).

- Open Preservation Foundation, Format Corpus, in: Github, <https://github.com/openpreserve/format-corpus> (06.03.2020).
- Open Preservation Foundation, <https://openpreservation.org/> (20.03.2020).
- Open Preservation Foundation, PAR Demo, <https://youtu.be/ELjbgksjBmU> (12.03.2020).
- OpenOffice.org XML, in: Wikipedia, https://en.wikipedia.org/wiki/OpenOffice.org_XML (03.03.2020).
- Pearson, David, AONS II. Continuing the Trend Towards Preservation Software ‚Nirvana‘, in: Proceedings of iPRES 2007 4th International Conference on Preservation of Digital Objects (2007), <https://pdfs.semanticscholar.org/163d/d56e1b160cd1b20dd5267d689f71d063760c.pdf> (17.03.2020).
- Präsentation der OPF: „Preservation Action Registries Workshop“, iPRES Event Workshop, September 2019, <http://parcore.org/presentations/PAR%20Presentation%20iPres%202019.pdf> (12.03.2020)
- Präsentation der OPF: „Preservation Action Registries Workshop“, DPC/JISC Workshop, 08.07.2019, <http://parcore.org/presentations/PAR%20Presentation%20York%202019.pdf> (12.03.2020).
- PREMIS Editorial Committee, PREMIS Data Dictionary for Preservation Metadata. Version 3.0 (2015), <http://www.loc.gov/standards/premis/v3/index.html> (20.01.2020).
- Preservica, PAR API, <https://preview.preservica.com/Registry/par/documentation.html> (12.03.2020).
- RDSS-PAR API, in: Github, <https://github.com/artefactual-labs/rdss-par/wiki> (12.03.2020).
- Rosenthal, David Stuart Holmes, Format Obsolescence: Scenarios (29.04.2007), <https://blog.dshr.org/2007/05/> (03.03.2020).
- Rosenthal, David Stuart Holmes, Format Obsolescence: the Prostate Cancer of Preservation (07.05.2007), <https://blog.dshr.org/2007/05/format-obsolescence-prostate-cancer-of.html> (03.03.2020).
- Rosenthal, David Stuart Holmes, The Half-Life of Digital Formats (24.11.2010), <https://blog.dshr.org/2010/11/half-life-of-digital-formats.html> (03.03.2020).
- Röthlisberger-Jourdan, Claire, Formaterkennung und Formatvalidierung. Theorie und Praxis, in: Keitel, Christian (Hrsg.), Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises „Archivierung von Unterlagen aus Digitalen Systemen“ und nestor-Workshop „Koordinierungsstellen“, Stuttgart 2013, S. 193–210.
- Röthlisberger-Jourdan, Claire, Sinn und Zweck der Formaterkennung und deren Grenzen. nestor-Praktikertag 2017, Kiel, 28.06.2017, https://kost-ceco.ch/cms/dl/6d2aa34e60f26b007d3eb3f0f86b061d/Formaterkennung_v1.0.ppsx (06.03.2020).

- Schweizerisches Bundesarchiv BAR, Partizipative Bewertung im BAR. Auswertung Pilot und Überführung Betrieb. Version 1.0 (2019), <https://www.bar.admin.ch/dam/bar/de/dokumente/diverses/Bericht%20BAR%20Pilot%20partizipative%20Bewertung.pdf.download.pdf/Bericht%20Auswertung%20Pilot%20partizipative%20Bewertung%20BAR.pdf> (21.02.2020).
- Schweizerisches Bundesarchiv BAR, Standards für die Archivierung digitaler Unterlagen. Archivtaugliche Dateiformate (2018), https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf (27.02.2020).
- Scott, Jason, Let's Just Solve the Problem Month (02.07.2012), <http://ascii.textfiles.com/archives/3645> (24.02.2020).
- Siegfried, <https://www.itforarchivists.com/> (06.03.2020).
- The National Archives, DROID, <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> (23.03.2020).
- The National Archives, File formats for transfer, <https://www.nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer/> (28.02.2020).
- The National Archives, PRONOM, The Technical Registry, <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx> (20.03.2020).
- Tilbury, Jon, A New Era in Collaboration in Digital Preservation Research (29.06.2018), <https://www.dpconline.org/blog/a-new-era-in-collaboration-in-digital-preservation-research> (11.03.2020).
- Tobegen, Michael, Entwicklung eines DIMAG-Bestandserhaltungsmoduls. Werkstattbericht, in: Naumann, Kai/Dahlmanns, Tim (Hrsg.), Von der Theorie zur Praxis: Bestandserhaltung digitaler Unterlagen. Tagungsdokumentation. Workshop des KLA-Ausschusses Digitale Archive, Bundesarchiv Koblenz, 7./8. November 2018, https://www.bundesarchiv.de/DE/Content/Downloads/KLA/tagungsdokumentation-bestandserhaltung-dig-unterlagen.pdf?__blob=publicationFile (24.02.2020).
- Todd, Malcom, Technology Watch Report. File formats for preservation (2009), http://www.dpconline.org/component/docman/doc_download/375-file-formats-for-preservation (07.02.2020).
- Träger, Christine/Wittmann, Daniel, Validierst du noch oder archivierst du schon... Der Bedarf eines Workflow-Managements in der Formatverifikation. 23. Tagung des Arbeitskreises AUdS, Prag, 12. und 13. März 2019, https://www.sg.ch/content/dam/sgch/kultur/staatsarchiv/dateien-stasg/camp-2-f/F_01_Traeger_AUdS_Praesentation_Formatverifikation.pdf (06.03.2020).
- Unified Digital Format Registry (UDFR), <http://www.udfr.org/> (24.02.2020).
- Worm, Peter, Nur ein Wunschkonzert? Hinweis auf eine Twitterumfrage und ihre Auswertung (2019), <https://archivamt.hypotheses.org/11487> (05.11.2019).

Gesprächsvermerke

Julia Bettinger, David Radlinger, Kai Naumann (alle Landesarchiv Baden-Württemberg), Stuttgart, 11.02.2020.

David Clipsham (PRONOM, The National Archives), Nicola Wissbrock (The National Archives), Kai Naumann (Landesarchiv Baden-Württemberg), Videochat, 17.02.2020.

Michael Tobegen (Staatsarchiv Hamburg), Telefonat, 27.02.2020.

Boris Kraut (Landesarchiv Baden-Württemberg), E-Mails vom 20.02.2020 und 10.03.2020.