

## Plädoyer für archivische Demut<sup>1</sup>

Georg Büchler, KOST  
Mai 2018

### Einleitung

Der *Call for Papers* für die 22. Jahrestagung des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen» warb im Besonderen um «Beiträge aus der Praxis der Archive, bei denen versucht wird, übergreifende Thesen und Schlussfolgerungen zu formulieren» und präzierte: «Dabei kann es sich gerne um kritische Rückblicke handeln.» Es ist ohne Zweifel sinnvoll, an einer solchen Tagung nicht nur Neues zu präsentieren, sondern auch Getanes zu reflektieren. Der vorliegende Beitrag möchte deshalb einen solchen kritischen Rückblick versuchen. Es ist ein Rückblick auf eine verbreitete, geradezu selbstverständliche Praxis, die weithin als *Good* oder *Best Practice* bezeichnet wird: Es ist in der digitalen Archivierung üblich, die zu archivierenden oder archivierten Unterlagen zu verändern. Das ist der Kern der postulierten Migrationsstrategie und in gewissen Fällen, ganz besonders bei der Archivierung aus Fachverfahren, kaum zu vermeiden<sup>2</sup>. Ein kritischer Blick auf Beispiele aus der Praxis enthüllt jedoch potenziell gefährliche Konsequenzen.

Zur Illustration seien hier drei Szenarien aus der Beratungspraxis der KOST vorgestellt, die vermutlich bei anderen Institutionen in ähnlicher Form ebenfalls vorkommen. Auf eine Analyse der in ihnen manifest werdenden Probleme folgen zum Schluss Elemente eines Lösungsansatzes.

---

<sup>1</sup> Schriftliche Fassung des Referats an der 22. Jahrestagung des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen» in Marburg, 7./8. März 2018. Ich danke den Tagungsteilnehmenden für die angeregte Diskussion, die sich auch im AUdS-Camp fortgesetzt hat und in den vorliegenden Artikel eingeflossen ist. Mein Dank geht ferner an meine Kollegin Claire Röthlisberger und meinen Kollegen Martin Kaiser für die gemeinsamen vorbereitenden Überlegungen und für ihre konkreten Rückmeldungen zur schriftlichen Fassung.

Sämtliche Weblinks wurden am 8. Mai 2018 zuletzt erfolgreich aufgerufen.

<sup>2</sup> Die Migrationsstrategie ist seit den neunziger Jahren in der digitalen Archivierung ein prominentes Thema. Einen guten frühen Überblick (mit älterer Literatur) bietet das Digital Preservation Testbed White Paper *Migration: Context and Current Status*, Den Haag 2001, gegenwärtig noch online zugänglich unter [http://kifri.fri.uniza.sk/~chochlik/diz\\_doc/sources/data\\_migr/Migration.pdf](http://kifri.fri.uniza.sk/~chochlik/diz_doc/sources/data_migr/Migration.pdf). In den letzten Jahren bekennen sich praktisch sämtliche staatlichen Archive im deutschsprachigen Raum zu dieser Strategie, siehe z.B. Schweizerisches Bundesarchiv (ed.), *Policy digitale Archivierung*, Bern 2009, [https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte\\_und\\_weisungen/policy\\_digitale\\_archivierung.pdf.download.pdf/policy\\_digitale\\_archivierung.pdf](https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/policy_digitale_archivierung.pdf.download.pdf/policy_digitale_archivierung.pdf).

## Drei Beispielszenarien

### *Szenario 1: Zeitnahe Konvertierung von Dokumenten in ein Archivformat*

Die meisten Archive, die bereits digitale Unterlagen übernehmen, verlangen deren Ablieferung in einem archivtauglichen Format<sup>3</sup>. Falls dieses nicht bereits das Ursprungsformat ist, impliziert diese Praxis, dass in der abliefernden Stelle eine Formatmigration stattfindet. Viele Archive empfehlen, diese möglichst früh im Lebenszyklus durchzuführen<sup>4</sup>. Die Absicht dahinter: Eine zeitnahe Konvertierung durch den Sachbearbeiter, beispielsweise im Moment der Finalisierung des Dokuments, ermöglicht es, nahe am Entstehungskontext zu bleiben, die für die Konvertierung notwendigen Ressourcen, in erster Linie die verwendeten Schriften, zugänglich zu haben und eine direkte Qualitätskontrolle zu gewährleisten.

In der Realität besteht jedoch abseits der archivischen Empfehlung kaum ein Bedürfnis nach einer solch frühen Konvertierung, zumal Office-Formaten schon lange nicht mehr so kurzlebig sind wie im ausgehenden 20. Jahrhundert<sup>5</sup>. So besteht die Gefahr, dass der archivischen Anforderung durch eine automatisierte Massenkonzertierung kurz vor der Ablieferung genüge getan wird. Das Archiv weiss in einem solchen Fall in der Regel weder, mit welchem Produkt, noch mit welchen Einstellungen gearbeitet wurde; der Entstehungskontext ist bereits verändert; die erhoffte Qualitätskontrolle findet höchstens stichprobenweise statt. Mit seiner gut gemeinten und theoretisch fundierten Formatstrategie zwingt das Archiv die abliefernden Stellen also zu einer potenziell verlustbehafteten Formatmigration in einem ungünstigen Moment. In diesem Fall wäre das Archiv besser beraten gewesen, das zum Zeitpunkt der Übernahme noch lesbare Ursprungsformat zu übernehmen.

### *Szenario 2: Mein Tool ist besser als deines.*

In den letzten Jahren ist die Toollandschaft im Bereich der digitalen Archivierung in erfreulicher Weise gewachsen<sup>6</sup>. Gewachsen sind auch die Akzeptanz und der Einsatz solcher Softwarewerkzeuge, beispielsweise im Bereich der Validierung von Dateiformaten<sup>7</sup>. Was dabei konkret geschieht, lässt sich auf eine einfache Formel herunterbrechen: Das Archiv benutzt Programme, um Dateien zu überprüfen, welche jemand anderes (in der Regel die abliefernde Stelle) mit anderen Programmen produziert hat. Wenn diese Prüfung nicht erfolgreich ist, wird das entsprechende Dokument zurückgewiesen.

Eine fehlgeschlagene Prüfung kann jedoch zwei verschiedene Ursachen haben: Entweder ist das Validierungstool korrekt und das erstellende Tool fehlerhaft – oder das

---

<sup>3</sup> Viele Archive publizieren dazu Listen von Archivformaten, z.B. Schweizerisches Bundesarchiv (ed.), *Archivtaugliche Dateiformate*, Bern 2018, [https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte\\_und\\_weisungen/archivtaugliche\\_dateiformate.1.pdf.download.pdf/archivtaugliche\\_dateiformate.pdf](https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf). Der *Katalog archivischer Dateiformate* der KOST (<https://www.kost-ceco.ch/wiki/whelp/KaD/index.php>) beruht auf den gleichen Überlegungen.

<sup>4</sup> So zuletzt nestor (ed.), *Die E-Akte in der Praxis. Ein Wegweiser zur Aussonderung*. Nestor-Materialien 20, Koblenz 2018, S. 11, <urn:nbn:de:0008-2018020827>.

<sup>5</sup> Beispielsweise kann die aktuelle Version von Microsoft Word Dokumente im Format Word 97-2003, die potenziell 20 Jahre alt sind, in der Regel noch einwandfrei anzeigen.

<sup>6</sup> Das «Community Owned digital Preservation Tool Registry» COPTR verzeichnet gegenwärtig 428 Tools zur digitalen Archivierung: [http://coptr.digipres.org/Main\\_Page](http://coptr.digipres.org/Main_Page).

<sup>7</sup> Hier sei in erster Linie auf die diversen Validatoren für PDF/A verwiesen; für eine aktuelle Übersicht siehe die KOST-Studie *PDF/A: Produktereview batchtauglicher PDF/A-Validatoren*, Bern 2018, [https://kost-ceco.ch/cms/index.php?pdfa\\_validatoren\\_de](https://kost-ceco.ch/cms/index.php?pdfa_validatoren_de).

erstellende Tool ist korrekt und das im Archiv verwendete Validierungstool ist fehlerhaft. Es gibt a priori keinen Grund, in diesem Fall standardmässig dem Validierungswerkzeug zu vertrauen, zumal erfahrungsgemäss keine Software vollständig fehlerfrei ist. Das sehr spezifische Fachwissen, um eine solche Situation richtig beurteilen und einordnen zu können, fehlt den Archiven in den meisten Fällen. Auf der Basis dieser sehr unsicheren Evidenz wird dann unter Umständen entweder die Datei im Archiv korrigiert, das heisst verändert, oder eine Ablieferung zurückgewiesen und damit der abliefernden Stelle beträchtlicher Aufwand verursacht (Kontrollen, Neuproduktionen) mit allen möglichen negativen Folgen.

### *Szenario 3: Archivierung von Daten aus Fachverfahren*

Zur Archivierung von Daten aus Fachverfahren empfiehlt die KOST das von ihr mitentwickelte SIARD-Format und die entsprechenden Tools<sup>8</sup>. Dieses Archivformat bewahrt die relationale bzw. Tabellennatur der Daten und ermöglicht so eine grosse Breite an künftigen Nutzungszielen. Das SIARD-Format wird international auch in diversen Archiven seit Jahren eingesetzt<sup>9</sup>.

Die Umstände der Entstehung dieser SIARD-Archivdatei entsprechen aber in der Regel keinesfalls gängiger *Best Practice*: Die zu archivierende Datei hat in der Realität niemals existiert. Sie wird in grossen zeitlichen Abständen (einmal jährlich oder seltener) ad hoc hergestellt, ohne Kontrolle des Archivs, durch einen Datenbankadministrator, der mit den Daten inhaltlich nicht vertraut ist. Jeder dieser Entstehungsumstände bietet Grund zur Besorgnis.

### *Weitere Beispiele*

Nur um zu illustrieren, dass das Problem nicht auf Einzelfälle beschränkt, sondern im Gegenteil relativ verbreitet ist, sei kurz auf zwei Beispiele verwiesen, die an Veranstaltungen im deutschsprachigen Raum in den letzten Jahren besprochen wurden.

Kai Naumann und Christoph Schmidt haben an der Jahrestagung 2017 des Arbeitskreises «Archivierung von Unterlagen aus digitalen Systemen» am Beispiel von Bildern dargelegt, dass der Informationsverlust, den eine klug ausgewählte Komprimierung verursacht, hinter dem bereits erfolgten Informationsverlust beim Digitalisieren weit zurücksteht, dass also fehlt, wer blind auf unkomprimierte Formate setzt, und vor allem, wer für die Ablieferung an das Archiv standardmässig in unkomprimierte Archivformate migrieren will<sup>10</sup>.

---

<sup>8</sup> Das SIARD-Format ist als eCH-0165 standardisiert, <https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165>, die aktuelle stabile Arbeitsversion ist zugänglich unter [https://kost-ceco.ch/cms/index.php?siard\\_de](https://kost-ceco.ch/cms/index.php?siard_de). Tools umfassen die SIARD Suite des Schweizerischen Bundesarchivs (<https://www.bar.admin.ch/bar/de/home/archivierung/tools---hilfsmittel/siard-suite.html>), die KOST-Tools KOST-Val (zur Validierung von SIARD-Dateien), csv2siard (zur Umwandlung von CSV-Dateien in SIARD) und SIARDexcerpt (für den Zugang zu den archivierten Informationen) sowie das DB Preservation Toolkit.

<sup>9</sup> Siehe <https://www.bar.admin.ch/bar/de/home/archivierung/tools---hilfsmittel/siard-suite.html>.

<sup>10</sup> Naumann, Kai; Schmidt, Christoph: «Chancen und Risiken des Einsatzes verlustbehafteter Bildkompression in der digitalen Archivierung», in: *Informationswissenschaft* 5,1 (2018), 59-71. Siehe zum gleichen Thema auch bereits KOST (ed.), *Datenkomprimierung bei Bild, Audio und Video*, 2009, [https://kost-ceco.ch/cms/index.php?compression\\_de](https://kost-ceco.ch/cms/index.php?compression_de).

Claire Röthlisberger hat 2016 am nestor-Praktikertag über die PDF/A-Validierung gesprochen und aufgezeigt, dass es nicht nur bessere und schlechtere Validatoren gibt<sup>11</sup>, sondern auch sozusagen einen Graubereich an den Rändern des Standards. Deshalb können sich Validatoren gelegentlich nach bestem Wissen und Gewissen uneinig sein und eine standardkonforme Datei als invalide identifizieren<sup>12</sup>. Soll dies für das Archiv ein Grund sein, in die Datei einzugreifen oder sie im Extremfall nicht zu übernehmen?

## Analyse und Lösungsansätze

In all diesen und ähnlichen Fällen kann nicht eigentlich von fehlerhaftem Handeln die Rede sein. Die geschilderten Praktiken beruhen auf richtigen Überlegungen und gültigen archivischen Prinzipien. Problematisch ist, dass dabei andere Prinzipien verletzt werden und keine bewusste Güterabwägung vorgenommen wird.

In Szenario 1 ist es vertretbar, wenn auch nicht zwingend, dass die Verantwortung für die Formatmigration bei der abliefernden Stelle liegt. Mit dieser Politik allein ist es allerdings nicht getan: Das Archiv muss sich darüber hinaus für deren konkrete Umsetzung interessieren.

In Szenario 2 ist die Verwendung von Tools keinesfalls in Frage zu stellen. Gefördert werden müssen jedoch unbedingt die Interpretation ihrer Ergebnisse und das Verständnis für ihre Limiten.

In Szenario 3 ist nicht zu kritisieren, dass die Natur der Daten und die zukünftigen Nutzungsziele wichtige Faktoren bei der Festlegung der Archivierungsstrategie sind. Kritisch betrachtet werden muss der Preis, der dafür bezahlt wird, das heisst, die konkreten Gefahren für die Integrität und Authentizität der archivierten Daten.

Dies alles bedeutet, dass es hier keine «richtige» Lösung geben kann. Weil die beschriebene Praxis nicht einfach falsch ist, ist auch nicht das Gegenteil einfach richtig. Die Probleme, und die gesamte digitale Archivierung, sind nicht schwarz-weiss.

Die Archivwissenschaft und unsere berufliche Tradition haben aber durchaus eine Leitlinie bereit, die im Titel dieses Beitrags etwas hochtrabend als archivische Tugend «Demut» bezeichnet wird. Grundlage der archivischen Profession und des archivischen Berufsethos ist, dass die Archivarin, der Archivar im Dienst der Archivalien steht. Es geht nicht um uns, es geht um die Archivalien. Auf dieser Grundlage zeichnen sich Auswege aus den skizzierten Dilemmata ab.

In Szenario 1 kann die Übernahme im Ursprungsformat, möglicherweise zusammen mit einer konvertierten Version, einen Ausweg darstellen. Ein solches Vorgehen ist traditionell verpönt. Grundlage dieser Einschätzung ist jedoch das längst veraltete Paradigma von der raschen Obsoleszenz von Dateiformaten, welche nur stabile Archivformate ein bisschen bremsen können. Inzwischen sind viele als nicht archivtauglich bezeichnete Formate seit mehr als zwei Jahrzehnten weit verbreitet und stabil und können nicht mehr nur mit Verweis auf ihre Kurzlebigkeit von der Archivierung ausgeschlossen werden. Auch sind viele vormals proprietäre Formate inzwischen gemeinfrei, offengelegt oder gut dokumentiert.

---

<sup>11</sup> Siehe dazu KOST (ed.), *PDF/A: Produktreview batchtauglicher PDF/A-Validatoren*, Bern 2017, [https://kost-ceco.ch/cms/index.php?pdfa\\_validatoren\\_de](https://kost-ceco.ch/cms/index.php?pdfa_validatoren_de).

<sup>12</sup> Röthlisberger, Claire: «KOST-Preservation-Prozess und die Erkenntnisse daraus», nestor-Praktikertag 2016, <http://www.langzeitarchivierung.de/Subsites/nestor/SharedDocs/Downloads/presentationen/2016PraktikertagRöthlisberger-Jourdan.pdf>.

Bei Szenario 2 lautet ein Lösungsansatz: «Kenne deine Tools!» Idealerweise kann das Archiv die wichtigsten Fehlermeldungen einordnen und ist mit der Materie ausreichend vertraut, um sich im richtigen Moment dem Urteil des Tools zu widersetzen. Ein anderer Lösungsansatz ist der Einsatz mehrerer Tools für die gleiche Analyse, was unter dem Begriff *triple-redundant designs* auch aus anderen Bereichen wie der Automobil- oder Luftfahrttechnik bekannt.

Zu Szenario 3 zeichnet sich keine einfache Lösung ab. Es ist nicht möglich, Datenbanken eins zu eins ins Archiv zu übernehmen und dort aktive Datenbanksysteme mit archivierten Daten zu betreiben. Eine Auswahl und verändernde Eingriffe sind also unabdingbar. Auch alternativ vorgeschlagene Lösungen wie die Veraktung von Fachverfahren erzeugen Archivalien, die so in der Verwaltungsrealität nie existiert haben, und ein PDF-Report aus einer Datenbank ist nicht weniger ein Ad-hoc-Produkt als eine SIARD-Datei. Die Problematik bedarf der weiteren Forschung, ein definitives Urteil steht noch aus. In der Zwischenzeit muss in diesem Bereich der digitalen Archivierung mit besonderer Sorgfalt und Augenmass gearbeitet werden.

Die Erleichterung darüber, in der digitalen Archivierung mit grossem Fachwissen und etwas Mut das eine oder andere erreicht zu haben, führt verständlicherweise leicht zu Euphorie. Aber darüber soll die archivische Demut nicht vergessen gehen. So wie wir heute mehr wissen als gestern (und deshalb gelegentlich zu forsch auf die gestrigen Arbeiten und Erkenntnisse zurückblicken), werden wir morgen mehr wissen als heute. Wir dürfen uns deshalb nicht zufrieden geben mit dem Erreichten, uns nicht verrennen in eine vermeintlich richtige Lösung, sondern müssen Fragen stellen, genau hinschauen, uns selber gegenüber kritisch bleiben – und demütig.