

Arbeitskreis „Elektronische Archivierung“ in der Vereinigung der Wirtschaftsarchivarinnen und Wirtschaftsarchivare (AKEA)

Bewertung und Pre-Ingest von kreativen Ablagen

Einleitung

Die Menge an digitalen Daten, die auf die Archive zukommen, wächst permanent. In diesem Zusammenhang wird die Übernahme von unstrukturierten Dateisammlungen zukünftig ein immer relevanter werdendes Thema und eine Herausforderung in den Archiven sein. Deshalb widmet sich diese Empfehlung dem Umgang mit Dateisammlungen, ihrer Bewertung und Aufbereitung sowie den Workflows beim Pre-Ingest und stellt in einer Tabelle hilfreiche Tools und Spezialwerkzeuge vor.

Die Archivar:innen sollen ermuntert werden, sich mit dem Thema aktiv auseinanderzusetzen. Auch wenn noch kein digitales Archiv oder Archivinformationssystem zur Verfügung steht, sollten sie bereits frühzeitig beginnen, Dateisammlungen von den abgebenden Stellen ihres Unternehmens zu übernehmen. Die Unternehmen arbeiten schon seit mehreren Jahren im hohen Maße digital, d.h. mit Office-Produkten, Datenbanken, elektronischen Akten, Online-Publikationen etc. Viele der entstehenden Dateien liegen dabei mehr oder minder gut strukturiert im Explorer Ordner. Die Archivar:innen können bei der Übernahme der digitalen Daten nicht warten, bis die gesetzlichen Aufbewahrungsfristen abgelaufen sind. Das kann für die Gewährleistung einer stringenten digitalen Überlieferung zu spät sein.

Dieses Paper soll daher dazu ermutigen, proaktiv und zeitnah auf die Abteilungen zuzugehen und noch vor Ablauf der gesetzlichen Aufbewahrungsfristen die digitalen Unterlagen zu übernehmen. Der Handlungsdruck erhöht sich stetig, deswegen lohnt es sich, mit der Übernahme kreativer Ablagen zum jetzigen Zeitpunkt anzufangen.

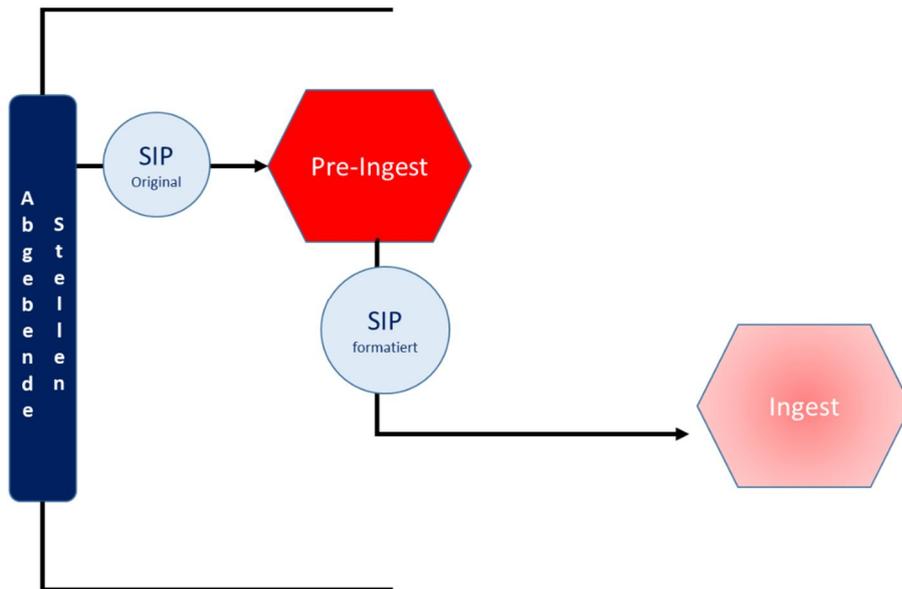


Abb. 1.: Ausschnitt aus dem OAIS-Modell, das sich auf die Workflows zwischen abgebenden Stellen, Pre-Ingest und Ingest beschränkt. Das SIP wird in ein „SIP Original“ und ein „SIP formatiert“ nach dem Pre-Ingest unterteilt.¹

Definition

Dateisammlungen werden synonym auch als „Fileablagen“, „unstrukturierte Daten“ oder „kreative Ablagen“ bezeichnet. Die Generaldirektion der Bayerischen Archive verwendete 2016 bei einer Tagung den Begriff: „kreative Ablage“², dem wir uns hier anschließen.

Wenn hier von kreativen Ablagen gesprochen wird, dann sind nicht gemeint:

- E-Akten in DMS- oder Vorgangsbearbeitungssystemen
- Fachverfahren
- Datenbanken
- E-Mail-Postfächer.

Eine kreative Ablage ist eine Sammlung von Einzeldateien, die von einem oder mehreren Bearbeiter:innen zur Erledigung einer oder mehrerer Aufgaben über einen bestimmten Zeitraum erstellt und nach individuellen Ordnungskriterien zusammengestellt wurden. Ihre Struktur ist ohne das Wissen des/der Produzent:in oft nicht verständlich bzw. nicht selbsterklärend. Die Logik der Ablage lässt sich nicht leicht

¹ Siehe dazu die Präsentation zum OAIS-Modell auf den Seiten des AKEA:

<https://www.wirtschaftsarchive.de/ueber-uns/arbeitskreise/fachliche-arbeitskreise/elektronische-archivierung/>.

² Aus: Sonderveröffentlichungen der Staatlichen Archive Bayerns Nr. 13 Kreative digitale Ablagen und die Archive, herausgegeben von Kai Naumann und Michael Puchta vom 22./23. November 2016

<https://www.bundesarchiv.de/DE/Content/Downloads/KLA/tagungsdokumentation-kreative-digitale-ablagen.pdf?blob=publicationFile>.

nachvollziehen. Die Dateien liegen auf einer Ebene und/oder hierarchisch in einer Verzeichnisstruktur vor. Es können in einer Dateisammlung unterschiedlichste Dateiformate enthalten sein. Meist liegen die Daten auf einer lokalen Festplatte, in Netzlaufwerken oder auf Gruppenlaufwerken. Es fehlen oftmals Angaben zum Entstehungszusammenhang. In manchen Fällen existieren neben den Fileablagen auch noch zusätzliche papierene Parallelüberlieferungen. Kurz: Eine vollständige Aktenbildung ist in der Fileablage zumeist nicht gegeben.

Herausforderungen bei der digitalen Bewertung

Die Bewertung von analogen und digitalen Unterlagen gehört zu den wichtigsten und anspruchsvollsten Aufgaben im Archivbereich. Qualitative Bewertungsmaßstäbe sind in der analogen und digitalen Welt (weitestgehend) identisch. Die Dokumentenart (Protokoll, Bericht, Vertrag etc.) ändert sich nicht. Die digitale Bewertung ist allerdings viel aufwendiger. Digitale Ordner lassen sich durch Menschen nicht so einfach „durchblättern“ und einfach forensisch untersuchen. Oft nehmen die aktenbildenden Stellen schon lange vor Abgabe an das Archiv selbst eine Bewertung vor, indem sie z.B. Teilbestände oder einzelne Dateien überschreiben oder löschen. Aktenbildende Stellen haben dabei oft ein falsches Verständnis von Datenschutz und dem „Recht auf Löschung“.³ Dies führt zum kontinuierlichen Verlust archivwürdiger Daten. Dateisammlungen enthalten zudem meistens unterschiedliche Dateiformate, die mit verschiedenen Programmen geöffnet werden müssen oder mitunter nicht in den Sammlungsbereich des Archivs fallen.

Eine weitere Schwierigkeit ist die Kurzlebigkeit von digitalen Daten und die Abhängigkeit von Speicherformaten. Archivar:innen scheitern oftmals an der Systemvielfalt und an den Berechtigungen. Hard- und Software entwickeln sich laufend weiter oder werden durch Migration abgelöst. Daten sind dadurch gegebenenfalls nicht mehr lesbar, gehen bei Migrationswechsel teilweise oder ganz verloren. Besonders problematisch ist die Ablage in personalisierten Umgebungen (z.B. persönliches Laufwerk oder Gruppenlaufwerke), zu denen Archivar:innen keinen Zugang haben oder deren Existenz nicht bekannt ist.

Angesichts des hohen Arbeitsaufwandes stellt sich die Frage, ob Archive auf die Übernahme kreativer Ablagen verzichten könnten. Da in vielen Unternehmen auch der Vorstand, die Geschäftsführung oder sonstige Entscheidungsträger eine Fileablage nutzen und in diesen potentiell archivwürdige Unterlagen speichern, sollte das Archiv diese Dateisammlungen jedoch sichten, bewerten und übernehmen.

³ Siehe u.a. Art. 89 Datenschutz-Grundverordnung (DSGVO) oder Art. 26 BayDSG. Diese regeln die Ausnahmen von der „Löschungspflicht“ aufgrund der im öffentlichen Interesse liegenden Archivzwecke.

Pre-Ingest: Workflows von der Analyse bis zum Ingest

Welche Schritte aber erfordert die Bewertung bzw. die Aufbereitung von kreativen Ablagen im Einzelnen? In der Arbeitsgruppe wurden verschiedenste Wünsche gesammelt und diskutiert, z.B. für die Phase der technischen und inhaltlichen Analyse und die Phasen der Nachbewertung, Übernahme und SIP-Formierung. Die Ergebnisse sollen hier nicht einzeln aufgeführt werden, sondern stattdessen das „Drei-Phasen-Modell“⁴ vorgestellt werden, welches drei Phasen unterscheidet:

- Analyse
- Nachbewertung und Formierung
- SIP-Erzeugung für das Archivinformationssystem

Die einzelnen Phasen wiederum bestehen aus verschiedenen Prozessschritten, wie die nachfolgenden Skizzen zeigen.

Phase 1: ANALYSE



Phase 2: NACHBEWERTUNG UND FORMIERUNG



⁴ Drei-Phasen-Modell nach Dr. Marco Birn (Kreisarchiv Reutlingen), Carolin Leitzbach (StA Ludwigsburg), Fabienne Metzger (Kreisarchiv Esslingen), Dr. Daniel Wilhelm (Universitätsarchiv Konstanz) <https://www.kultur-machen.de/dateisammlungen>.

Phase 3: SIP-ERZEUGUNG FÜR DAS ARCHIVINFORMATIONSSYSTEM



Abb. 2: Musterworkflows des Drei-Phasen-Modells zur Bearbeitung von Dateisammlungen, Grafiken übernommen von Dr. Marco Birn.

Eine ausführliche Beschreibung der Arbeitsschritte der einzelnen Phasen findet sich in der Handreichung „Musterworkflow zur Bearbeitung von Dateisammlungen“ von Dr. Marco Birn, Carolin Leitzbach M.A., Fabienne Metzger M.A. und Dr. Daniel Wilhelm (vgl. Fußnote 4).

An diesem Drei-Phasen-Modell können sich die Archivar:innen gut orientieren, um erste Schritte beim Umgang mit Dateisammlungen in die Wege zu leiten. Selbst wenn sie noch kein digitales Archiv oder ein Archivinformationssystem im Einsatz haben, kann vieles bereits vorbereitet werden.

Der Vollständigkeit halber soll noch erwähnt werden, dass es in der Archivwelt auch Stimmen gibt, die diese „Bearbeitung“ kritisch sehen, denn durch die oben erwähnten Schritte werden Teile der Überlieferung verfälscht. Hinzu kommt der hohe Arbeitsaufwand vor allem durch den Einsatz von Personal.

Übersicht über die Pre-Ingest-Tools (siehe Anlage Excel-Liste)

Die in der Übersicht aufgeführten Tools sind jeweils entweder den Mitgliedern des Arbeitskreises aus der eigenen Pre-Ingest-Praxis bekannt oder im Rahmen von archivischen Fortbildungen und Fachvorträgen vorgestellt und empfohlen worden.

Die Tabelle gibt eine Übersicht über Tools, die für den Pre-Ingest eingesetzt werden können. Berücksichtigt sind hierbei verschiedene Arten. Neben den beiden speziell für Archive und deren Anforderungen entwickelten Tools (COMO, docuteam packer) sind Anwendungen in die Übersicht aufgenommen, die zwar nicht eigens für Archive entwickelt wurden, jedoch hilfreich für mehrere Pre-Ingest-Schritte sind (TreeSize, Dateilisten Schreiber, Total/Free Commander, iFinder). In einer separaten Spalte werden daneben Tools genannt, die die einzelnen Schritte beim Pre-Ingest unterstützen (z.B. KostVal). Zum Vergleich wurde in die Übersicht daneben auch das Programm IngestList aufgenommen, das allerdings nicht mehr weiterentwickelt wird und nur für DIMAG-Nutzer:innen geeignet ist.

Für jedes Tool wird zunächst eine Kurzbeschreibung gegeben. In den darunter liegenden Zeilen sind in Anlehnung an Arbeitsschritte beim Pre-Ingest verschiedene Funktionalitäten genannt. Es wird für jedes Tool angegeben, ob es diese Funktionalität besitzt und damit der entsprechende Pre-Ingest-Schritt mit Hilfe der Anwendung ausgeführt werden kann.

Sofern ein Tool von Mitgliedern des Arbeitskreises bereits verwendet wurde oder wird, sind ferner die Erfahrungen hinsichtlich der Handhabung und der Performance in der Tabelle kurz zusammengefasst.

Der Übergang zwischen Pre-Ingest und Ingest ist mitunter fließend, sodass eine klare Trennung zwischen Pre-Ingest-Tool und Ingest-Tool nicht immer klar möglich ist. Beispielsweise werden einzelne Schritte, die in Regel dem Pre-Ingest zugeordnet werden, beim Anbieter docuteam nicht durch den in dieser Übersicht als Pre-Ingest-Tool beschriebenen docuteam packer ausgeführt, sondern erst beim Ingest über den docuteam feeder. Auch bei COMO sind Überschneidungen zwischen Pre-Ingest und Ingest feststellbar. Aus diesem Grund werden in der Übersicht bei den Tools COMO und docuteam packer bei einzelnen Pre-Ingest-Schritten weitere Anwendungen der beiden Anbieter startext bzw. docuteam genannt.

AKEA-Arbeitsgruppe (Stand 2023)

Frank Becker (evonik); Bettina Hasselbring (Bayerischer Rundfunk); Oliver Laux (debeka); Sandra Leibner (Bayerischer Rundfunk); Benjamin Machauer (BMW); Elke Pfnür (UniCredit Bank AG); Edith Plöthner (Genossenschaftsverband Bayern e.V.); Sabine Rittner (Bayerischer Rundfunk); Matthias Weber (ECB); Susanne Witschaß-Bayer (BDI)