

Überlegungen zur Archivierung und langfristigen Nutzbarhaltung von Fachapplikationen

Tobias Wildi, Fachhochschule Graubünden, tobias.wildi@fhgr.ch

9.7.2024

Inhalt

Neue Überlegungen zur Archivierung und Nutzbarhaltung von Fachapplikationen	1
Einleitung	2
Eine Kombination aus Daten- und Softwarearchivierung	6
Bewertung	7
Ablieferung	9
Datenzentrierte Ablieferung	9
Objektzentrierte Ablieferung	10
Applikationszentrierte Ablieferung	11
Zugang und Vermittlung	13
Linked Data als applikationsunabhängiges Nutzungsformat	13
Publikation von SIARD als SIARD-RDF auf dem LINDAS-Portal	15
Fazit	16
Literatur	16

Einleitung

Wie und wo schlagen sich die Entscheidungen und Handlungen aus Verwaltungsprozessen nieder?¹ Und wie werden diese Prozesse im Archiv dokumentiert und dauerhaft nachvollziehbar gehalten? Es sind keine grundsätzlich neuen Fragen, die dieser Artikel stellt. Angesichts der Digitalisierung und Automatisierung von Verwaltungsprozessen erhalten sie jedoch neue Aktualität. Der traditionelle und oft auch in den Archivgesetzen verankerte Weg sieht vor, im Archiv mehr oder weniger passiv abzuwarten, bis die Verwaltung Unterlagen abliefern. Das Archiv soll nicht in die aktive Phase des Lebenszyklus eingreifen und wenn, dann höchstens in beratender Funktion im Bereich des Records- und Informationsmanagements.

Dieser Ansatz ist in einer vernetzten digitalen Verwaltungsrealität schwierig aufrecht zu halten. Denn ohne kontextualisierende Metadaten und Strukturinformationen, die bereits bei der Unterlagenproduktion angelegt werden, ist die Bewertung und Übernahme digitaler Ablieferungen kaum bewältigbar. Das hat sich in der Vergangenheit beispielsweise bei den Übernahmen grosser Dateiablagen gezeigt, ein Vorgang, der aufgrund fehlender Metadaten nur schwer skaliert und automatisiert werden kann (Puchta & Naumann, 2017). In den letzten 25 Jahren haben die Archive auf die Herausforderung reagiert, indem sie den Fokus primär auf unstrukturierte Unterlagen gelegt haben und mit den Methoden des Informations- und Records Managements für prozessorientierte Ablagestrukturen in den Verwaltungen gesorgt haben. Ein erfolgreiches Beispiele dafür ist in der Schweiz die Einführung der elektronischen Geschäftsverwaltung in der Bundesverwaltung (GEVER),² auch viele Kantone haben ähnliche Programme bereits abgeschlossen oder setzen sie noch um.³ In der archivischen Fachdiskussion bisher viel weniger thematisiert wurde die Übernahme strukturierter Daten aus Fachapplikationen. Fachapplikationen, in Deutschland auch Fachverfahren genannt, sind in der Regel datenbankgestützte Systeme, die die strukturierte Abarbeitung standardisierter Verwaltungsprozesse unterstützen, respektive diese erst ermöglichen. Der funktionale Umfang solcher Systeme weist eine grosse Spannweite auf und reicht von kleinen Register-Datenbanken in Gemeindeverwaltungen bis zu SAP-gestützten Grosssysteme auf Stufe Bundesverwaltung. Diskussionsansätze zur Bewertung und Archivierung von Fachapplikationen existieren etwa in Deutschland im VdA (VdA-Arbeitskreis «Archivische Bewertung», 2014) oder in der Schweiz in der VSA-Arbeitsgruppe «Bewertung» im laufenden Projekt zur Erarbeitung einer «Checkliste zur archivischen Bewertung von Fachapplikationen».⁴ Darüber hinaus konzentriert sich die Fachliteratur aber weitgehend auf konkrete Erfahrungsberichte pionierhafter Datenübernahmen aus Altanwendungen. Eine Diskussion übergreifender Fragen findet erst in Ansätzen statt. Beispielsweise widmet sich die Ausgabe 01/2023 der Zeitschrift «Archiv theorie & praxis» dem

¹ Der vorliegende Text entstand im Rahmen eines Projekts an der Fachhochschule Graubünden zur Archivierung von Datenbanken mit Hilfe von Linked Data. Das Projekt wurde finanziert durch das Schweizerische Bundesarchiv. Der Autor bedankt sich für die Unterstützung.

² <https://www.bk.admin.ch/bk/de/home/dokumentation/gever-bund.html>

³ Etwa der Kanton Bern:

<https://www.staatsarchiv.sta.be.ch/de/start/fuer-behoerden/programm-dga.html>

⁴ <https://vsa-aas.ch/verein/arbeitsgruppen/bewertung/>

Schwerpunktthema “Fachverfahren als Aspekt moderner Arbeit”.⁵ Weitere generalisierbare Erkenntnisse finden sich im 2023 publizierten nestor-Standard “Archivierung von Studierendendaten aus Fachverfahren” (nestor 2023).

Ein charakteristisches Merkmal von Fachapplikationen ist die oft erstaunliche lange Einsatzdauer, die in vielen Fällen über ein Jahrzehnt beträgt. Die Applikationen werden während dieser Zeit weiterentwickelt, verändert und an neue Bedürfnisse angepasst. Die Archivierung der Daten kann nicht einfach aufgeschoben werden, bis eine Fachapplikation als “End of Life” deklariert wird und durch eine nächste Generation ersetzt wird. Erstens ist das Risiko nicht dokumentierter Änderungen an den Daten gross, selbst wenn die bearbeiteten Fälle längst abgeschlossen sind. Relationale Datenbanksysteme sind grundsätzlich nicht darauf ausgelegt, die Historie von Datenänderungen zu protokollieren. Zweitens stehen die nicht archivierten Daten im Archiv entsprechend auch nicht zur Konsultation zur Verfügung. Der vorliegende Artikel adressiert diese Problematik. Er verfolgt das Ziel, archivische Fragestellungen im Kontext der Architektur, Bewertung, Ablieferung und Vermittlung von Fachapplikationen zu erörtern und zu systematisieren.

Folgendes Beispiel illustriert die Problematik auf anschauliche Weise. 2017/2018 führte die schweizerische KOST (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen) zusammen mit einigen Staatsarchiven das Projekt “Archivierung der staatlichen Finanzdaten” durch.⁶ Der im Anschluss veröffentlichte Bericht (Ducry et al. 2018) nennt zentrale Herausforderungen, die nicht nur für Finanzdaten, sondern generell für die Archivierung aus allen Typen von Fachapplikationen gelten. Der Bericht stellt fest, dass in den Verwaltungen der Kantone vorrangig die zentral erstellten Endprodukte der Finanzprozesse archiviert werden. Konkret heisst es: “Es werden in erster Linie die Hauptprodukte der Finanzämter, d.h. die Staatsrechnung und das Budget, archiviert. Diese werden oftmals als Publikationen und nicht als Archivadokumente behandelt. Es scheint gängige Praxis zu sein, dass die Ablieferung zentral durch das Finanzamt erfolgt und die Finanzunterlagen der Verwaltungseinheiten in der Regel entsorgt werden.” (Ducry et al., 2018, S. 31). Verwaltungsprozesse involvieren oft mehrere datenproduzierende Organisationseinheiten und Daten liegen in mehreren Aggregationsstufen vor, von den Rohdaten bis zu den Endauswertungen. Beides muss bei der Bewertung und Ablieferung der Daten berücksichtigt werden und die ausschliessliche Fokussierung auf die Endergebnisse greift zu kurz.

Weiter hebt der Bericht hervor, dass die Endprodukte der Finanzämter in der Regel als PDF veröffentlicht und in diesem Format an die Archive übergeben werden. Diese Praxis führt zu Einschränkungen für zukünftige Datenverarbeitung und -analyse: “Bei der Archivierung in Form von unstrukturierten Daten entfällt die Möglichkeit der freien Auswertung der Daten, und die Verwendung von Produkten wird aufgrund der starren Form von PDF/A-Dateien eingeschränkt. Eine maschinelle Analyse der in PDF/A-Dateien enthaltenen Daten ist zwar denkbar, aber sehr kostspielig” (Ducry et al., 2018, p. 29). Die bei der Ablieferung gewählten Daten- und Dateiformate

⁵ Siehe

<https://www.archive.nrw.de/sites/default/files/media/files/Archivar20231InternetmitAnzeigen.pdf>

⁶ https://kost-ceco.ch/cms/17-036_de.html

haben Implikationen für die spätere Vermittlung und Nachnutzung der Daten. In der aktuellen Diskussion wird die traditionelle Auffassung, dass Datenbankapplikationen ausschliesslich als Hilfsmittel zur Generierung von Endprodukten dienen und daher keinen Archivwert besitzen, zunehmend als veraltet angesehen.

Doch welche innovativen Ansätze werden heute diskutiert, die die Übernahme und langfristige Nutzbarhaltung von Daten aus Fachapplikationen unterstützen könnte? Neue Ideen finden sich beispielsweise im Bereich des Lifecycle-Managements. Der Holländer Geert-Jan van Bussel schlägt in seinem Paper "The Theoretical Framework for the 'Archive-As-Is'" (Van Bussel, 2017) ein holistisches Lifecycle-Konzept vor, das aus insgesamt fünf Komponenten besteht. "Archive-As-Is" berücksichtigt neben den Records und den archivischen Prinzipien auch stark die zukünftige Nutzung. Zudem rückt das Modell auch Fragen bezüglich Umsetzung und Einfluss der Organisationsform ins Blickfeld des Frameworks, welches mehr will, als nur den Informations-Lebenszyklus neu zu ordnen. Dieser Ansatz soll zu einer ganzheitlichen Betrachtung von Fachapplikationen über sämtliche Lebensphasen hinweg führen.

Ein zweiter Ansatz unter der Bezeichnung "Archiving by Design" und wurde durch das niederländische Nationalarchiv entwickelt (Hooft, 2023).⁷ Hierbei wird ein neuer Ansatz propagiert, der bereits beim Design von Fachapplikationen einsetzt. Die langfristige und nachhaltige Nutzung und Verfügbarkeit von Informationsressourcen soll als Funktionalität bei der Konzeption und Umsetzung neuer Prozesse und Applikationen systematisch mitbedacht werden. Das eröffnet die Chance, in einer konzeptionellen Phase den Benutzerbedürfnissen bezüglich langfristiger Nutzung von Datenbeständen mehr Gewicht zu geben. Heute erschwert für die Verwaltung beispielsweise die systemische Trennung von Produktiv- und Archivsystemen die Recherche und Nutzung im Rahmen des Wissensmanagements.

Beide genannten Ansätze beschränken sich nicht bloss auf die Revidierung bekannter Lebensphasen-Modelle, sondern berücksichtigen durchaus auch sich verändernde Arbeitspraktiken und Organisationsformen der Verwaltung wie Kollaboration oder Ko-Kreation. Zudem legen sie die konzeptuellen Grundlagen, die nachhaltige Zugänglichkeit und Nutzung von Informationsressourcen zu stärken.

Es bietet sich die Gelegenheit, auch geleitet durch diese neuen Ansätze, einen systematischen Blick auf die Archivierung von Fachapplikationen zu legen und zwar mit folgenden Fragen:

- **Was archivieren?** Welche Aspekte von Fachapplikationen sind überhaupt archivwürdig? Die naheliegende Antwort lautet, dass es primär die Daten oder aggregierte Auszüge (Reports, Views) sind. Bisher wenig diskutiert wird der Wert der Applikationen als Software-Artefakte. Die Applikationen dokumentieren mit ihrem technischen Stand und ihren Mensch-Maschinen-Schnittstellen die Verwaltungspraktiken einer bestimmten Zeit. Gerade die Veränderungen der Benutzeroberflächen von textbasierten Terminals über erste grafische Oberflächen bis zu den heutigen webbasierten Systemen zeigen die fundamentalen Änderungen, wie Mensch und Maschine in der Verwaltung interagieren.

⁷ <https://www.nationaalarchief.nl/en/archive/knowledge-base/archiving-by-design>

Sucht man heute in einem öffentlichen Archiv nach Softwarecode oder Screenshots von Fachapplikationen, so wird man kaum fündig.

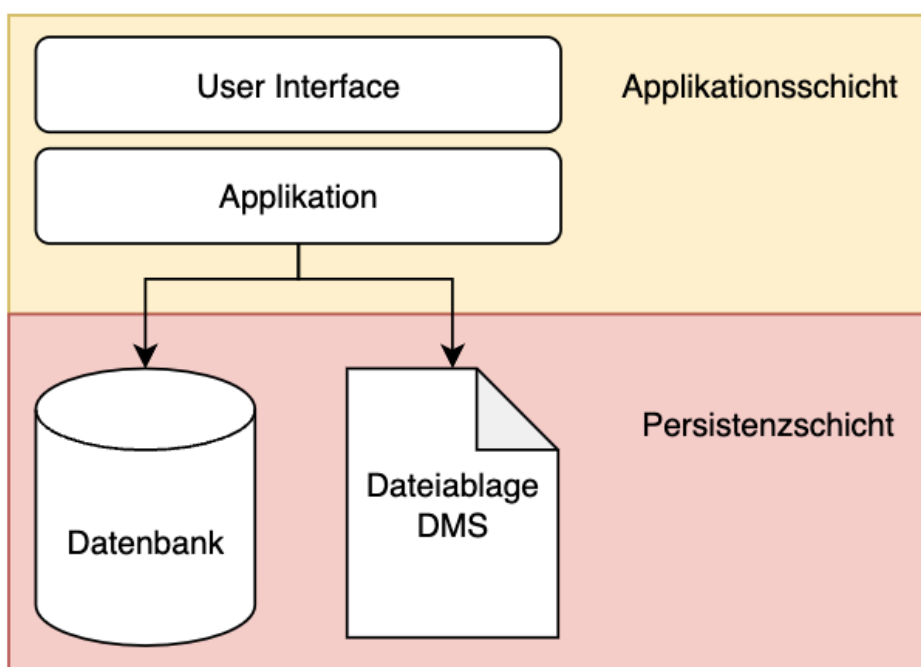
- **Wie archivieren?** Die archivwürdigen Daten sollen im Archiv nutz- und auswertbar gehalten werden, unabhängig von den Applikationen, mit denen sie erzeugt wurden. Für eine applikationsneutrale Archivierung braucht es flexible Datenmodelle und standardisierte Schnittstellen, über die die Daten der Verwaltung und der Öffentlichkeit zur Verfügung gestellt werden können.
- **Wie zugänglich machen?** Datenbanken werden nach heutiger Praxis in Formaten archiviert, die sich wenig für die Benutzung im Lesesaal eignen. Welche technischen Möglichkeiten gibt es also, gerade auch im Bereich von Linked Data, um die Zugänglichkeit zu verbessern?

Diesen drei Fragen geht der vorliegende Artikel nach. In einem ersten Teil werden die Architektur und der Aufbau von Fachapplikationen thematisiert und grundsätzliche Strategien zur Archivierung von Fachapplikationen erörtert. Im zweiten Teil wird die Bewertung von Fachapplikationen auf der Basis der Geschäftsprozesse und Aufgaben beschrieben. Im dritten Teil steht die Durchführung von Ablieferungen, respektive Übernahme von Daten ins Archiv im Vordergrund. Im fünften Teil wird das Potential von Linked Data und semantischen Netzen ausgelotet, um den Archivbenutzer:innen Datenbestände aus Fachapplikationen in einem applikationsneutralen Format zur Verfügung zu stellen und mit weiteren Datenquellen in Beziehung zu setzen.

Eine Kombination aus Daten- und Softwarearchivierung

Neben Records Management- und Geschäftsverwaltungssystemen sind Fachapplikationen die wichtigsten datenproduzierenden Systeme der öffentlichen Verwaltung. Die Architektur von Fachapplikationen unterteilt sich grob in eine Applikations- und eine Persistenzschicht, letztere umfasst typischerweise auch eine Datenbank. Die Archivierung von Fachapplikationen stellt besondere Anforderungen, weil sich mit der alleinigen Archivierung der Datenbank die Geschäftslogik oft noch nicht nachvollziehen lässt.

Typischerweise handelt es sich bei Fachapplikationen um datenbankbasierte Systeme, welche zur Abbildung standardisierter Prozesse verwendet werden. Teilweise sind die Datenbanken noch erweitert um Dateiablagen. Die grobe funktionale Architektur unterteilt sich in folgende Elemente:



Die Bereiche "Applikation" und "User Interface" umfassen die Programmlogik und die Benutzer- und Maschinenschnittstellen. Die "Datenbank" ist darauf ausgerichtet, strukturierte Daten effizient und zuverlässig zu speichern. Und schliesslich ist die "Dateiablage/DMS" auf die Speicherung binärer Objekte (Dateien) ausgelegt. Unterschieden wird also grob zwischen einer Applikationsschicht und einer Persistenzschicht.

In der technischen Umsetzung sind die drei Funktionen je nach Fall systemisch voneinander getrennt oder fallen zusammen. Beispielsweise sind moderne Datenbanksysteme durchaus in der Lage, Dateien als "Binary Large Objects" (BLOB) zu speichern. Viele Fachapplikationen speichern Dateien aber direkt im Dateisystem oder in einem Dokumentenmanagementsystem (DMS). Die funktionale Architektur sagt auch nichts aus über den Betriebsmodus der Fachapplikation, sei es lokal ("on

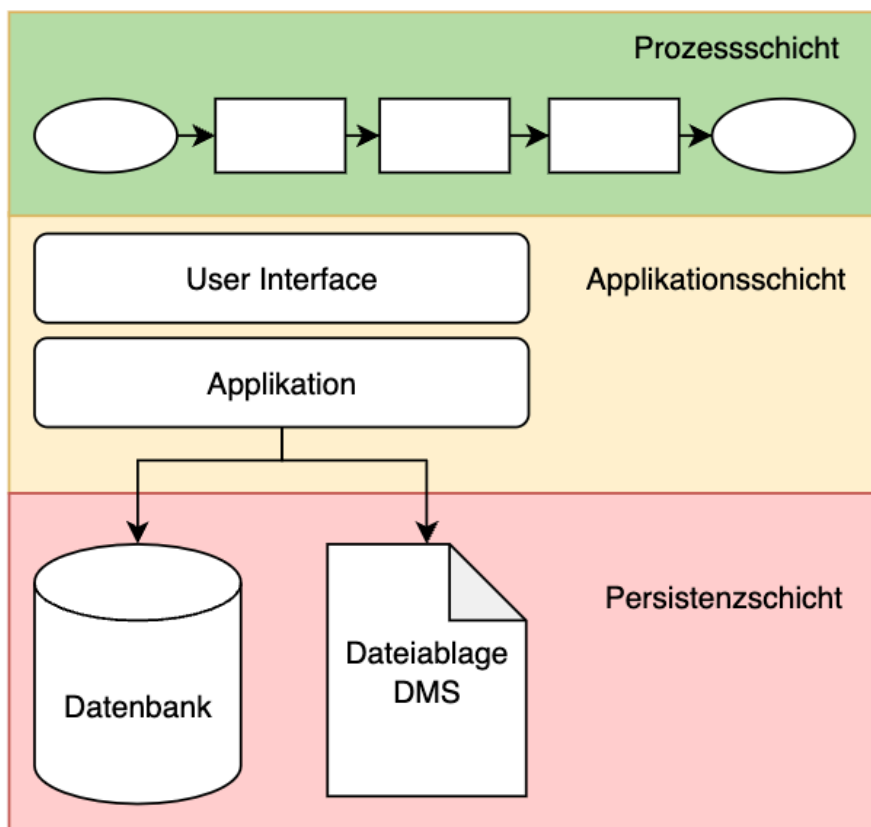
premises”) oder als Dienst aus der Cloud (“Software as a Service”). Die unterschiedlichen Betriebsmodelle sind für die vorliegenden Überlegungen denn auch nicht weiter von Bedeutung. Für die Bewertung und Archivierung ist es indessen essentiell, dass alle funktionalen Bereiche und nicht nur die Datenbank betrachtet werden. Die Archivierung betrifft nicht nur die Rohdaten in der Persistenzschicht, sondern potenziell auch die Programmlogik und die Schnittstellen in der Applikationsschicht. Denn je nach den langfristig zu erhaltenden Eigenschaften, vorgegeben durch das Bewertungsprofil und den “Preservation Intent”, gilt es mit einer geeigneten Archivstrategie unterschiedliche Elemente langfristig zu erhalten. Wir sind konfrontiert mit einer komplexen und mehrschichtigen Kombination aus Daten- und Softwarearchivierung.

Bewertung

Nur in den wenigsten Fällen wird eine Fachapplikation im archivischen Bewertungsprozess global als archivwürdig oder nicht archivwürdig eingestuft. Meist bestimmt ein Archiv eine Auswahl an Elementen, die übernommen werden sollen. Dazu werden die Prozesse und Aufgaben analysiert, welche durch die Fachapplikationen abgebildet werden. Auf dieser Basis ist dann die Erstellung eines systematischen Bewertungsprofils und die Benennung der archivwürdigen Daten möglich.

Bei der Bewertung von Fachapplikationen stellt sich die Frage, welche ihrer Eigenschaften und Daten als archivwürdig eingestuft werden. Die Bewertung gründet sich primär auf die Prozesse und Aufgaben, die von einer Fachapplikation abgedeckt oder erfüllt werden. Diese Aufgaben basieren ihrerseits auf einem Mandat oder einem Auftrag. Im Kontext öffentlicher Körperschaften beruht dies auf einer gesetzlichen Grundlage, während private Organisationen in der Regel die Freiheit haben, sich einen eigenen Auftrag zu geben.

Die Bewertung erfolgt also unabhängig von der oben beschriebenen Applikations- und Persistenzschicht und wird auf der Ebene der Aufgaben und Geschäftsprozesse einer Organisationseinheit durchgeführt (siehe Niu, 2013, insb. Abschnitt “Appraisal Criteria”). Das Schweizerische Bundesarchiv schreibt in seinem “Gesamtkonzept für die Bewertung im Bundesarchiv”: “Innerhalb des Geltungsbereiches des BGA [Bundesgesetz über die Archivierung] soll abgebildet werden, wie Kompetenzen und Aufgaben wahrgenommen werden. Dabei ist die Bewertung nach dem Provenienzprinzip organisiert: Es werden die gesetzlichen Kompetenzen, Aufgaben und Organisation der einzelnen Stellen ermittelt.” (Schweizerisches Bundesarchiv, 2010). Die Operationalisierung, also die Bestimmung, was in welcher Form und zu welchem Zeitpunkt abgeliefert werden soll, erfolgt anschliessend in der Persistenz- und Applikationsschicht. Demnach muss das Architekturmodell um eine zusätzliche Ebene erweitert werden:



Kaum je läuft ein Bewertungsentscheid darauf hinaus, eine Fachapplikation “als Ganzes” zu archivieren, also sowohl die kompletten Daten in der Persistenzschicht als auch das Softwaresystem der Applikationsschicht zu erhalten. Um das Dokumentationsziel zu erfüllen, genügen meist schon aggregierte oder nach bestimmten Kriterien ausgewählte Daten. Dieses Vorgehen lässt sich beispielsweise an den Bewertungsentscheiden der schweizerischen Bundesverwaltung nachzeichnen, welche das Schweizerische Bundesarchiv regelmässig im Überblick “Bewertungen zu Inhalten aus Informationssystemen des Bundes” publiziert.⁸ Als archivwürdig eingestuft werden grundsätzlich Fachapplikationen, deren Abläufe und Prozesse langfristig nachvollziehbar bleiben sollen. Nur bei einigen wenigen Datenbanken der Bundesverwaltung bedeutet dies die “integrale Archivierung der Daten (und Funktionalitäten) der Applikation” oder die “integrale Archivierung der Inhalte, Funktionalitäten und Dokumentation”, also tatsächlich soweit möglich die gesamte Persistenz- und Applikationsschicht. Bei anderen Datenbanken fokussiert die Bewertung auf eine Auswahl von Fällen. Die Auswahl hat, je nachdem, mittels Sampling, Stichprobe oder manueller Auswahl von Einzelfällen zu erfolgen. Beispiele für solche Entscheide lauten etwa: “Archivierung der Daten in Auswahl (Selektion)” oder “Sampling, 1%

⁸ Bewertungen zu Inhalten aus Informationssystemen der schweizerischen Bundesverwaltung, Stand 2024, Link zur Excel-Datei ganz unten auf folgender Seite: <https://www.bar.admin.ch/bar/de/home/informationsmanagement/archivwuerdigkeit/bewertungen-tscheide.html>

der Dossiers". Als nicht archivwürdig eingestuft werden in der Regel Daten zur Benutzerverwaltung, technische Log- und Metadaten oder Hilfsdatenbanken.

Zusammengenommen finden sich drei unterschiedliche Kategorien, wie bewertet und dann abgeliefert werden soll:

- **Datenzentriert:** Bei diesen Ablieferungen werden Daten aus einer Datenbank exportiert und ans Archiv übermittelt. Dabei kann es sich um einen integralen Datenbank-Export handeln oder um eine Auswahl, die beispielsweise mittels einer View erstellt wird.
- **Fall- oder objektzentriert:** Das Archiv beschränkt sich auf die Übernahme einer Auswahl von Fällen aus der Fachapplikation, denn nicht immer braucht es die gesamte (umfangreiche) Datenbank, um das Funktionieren und die Entscheidungsprozesse einer Behörde nachvollziehen zu können. Bei diesen Ablieferungen werden Fälle nach vorgegebenen Kriterien ausgewählt und aus der Fachapplikation exportiert.
- **Applikationszentriert:** Um die archivwürdigen Prozesse nachzuvollziehen, reichen in manchen Fällen die Daten alleine nicht aus, sondern es müssen auch einige oder alle Funktionalitäten der Applikation mitarchiviert werden. Das Archiv übernimmt hier also nicht nur Daten, sondern auch Software.

Ablieferung

Wenn ein Bewertungsentscheid vorliegt, muss eine konkrete Vorgehensweisen entwickelt werden, um die Daten oder auch die Software ins Archiv zu übernehmen. Entsprechend der Art des Bewertungsentscheids werden im folgenden Kapitel drei Ansätze für die Übernahme diskutiert: Erstens ein datenzentriertes Vorgehen, wo es um die Archivierung ganzer Datenbanken geht. Zweitens ein objektzentriertes Vorgehen, wo alle zu einem Geschäftsfall gehörenden Daten abgeliefert werden. Und drittens ein applikationszentriertes Vorgehen, wo die Fachapplikation als Softwareartefakt im Zentrum des Interesses steht.

Datenzentrierte Ablieferung

Bei der datenzentrierten Ablieferung werden in der Persistenzschicht Daten für die Archivierung ausgewählt, exportiert und abgeliefert. Mit dem XML-basierten SIARD-Format des Schweizerischen Bundesarchivs besteht die Möglichkeit, relationale Datenbanken in applikationsneutraler Form zu archivieren.⁹ SIARD ist mittlerweile ein international anerkannter Standard, der Verbreitung in

⁹ SIARD Formatbeschreibung V2.2 (2021):

https://www.bar.admin.ch/dam/bar/de/dokumente/kundeninformation/siard_formatbeschreibung.pdf.download.pdf/siard_formatbeschreibung.pdf

zahlreichen Archiven gefunden hat.¹⁰ Einfach strukturierte Fachapplikationen, beispielsweise Register, werden oft auch als CSV- oder Excel-Tabelle ins Archiv übernommen.

Der Vorteil der datenzentrierten Ablieferung liegt in der vergleichsweise einfachen technischen Umsetzung. Die Daten werden entweder über eine Schnittstelle aus dem Datenbankserver exportiert und in ein Archivformat umgewandelt oder die Archivierungssoftware verbindet sich direkt mit dem Datenbankserver und exportiert die Daten, so dass datenbankseitig keine Vorkehrungen getroffen werden müssen. Für die Umwandlung in das SIARD-Format existiert frei verfügbare Software wie die "SIARD Suite" des Schweizerischen Bundesarchivs¹¹ oder das "Database Preservation Toolkit".¹² Beide Werkzeuge sind in der Lage, Daten von unterschiedlichen Datenbankservern abzugreifen. Die datenzentrierte Ablieferung hat allerdings auch Nachteile. Für die Konsultation einzelner Datenbankeinträge muss im Archiv jeweils die gesamte Datenbank aus dem Repository geladen werden. Dabei kann es sich um erhebliche Datenmengen handeln, vor allem wenn Fachapplikationen während vieler Jahre in Betrieb standen und zusätzlich noch Binärobjekte (Dateien) gespeichert wurden. Zudem kann aus der Datenbankstruktur in der Persistenzschicht in vielen Fällen nicht auf die Programmlogik in der Applikationsschicht zurückgeschlossen werden. Aus der Datenbankstruktur erschliesst sich noch nicht, wie die Verwaltung gearbeitet hat. Den Archivbenutzer:innen wird die Auswertung der Datenbank Schwierigkeiten bereiten, wenn sie nicht über zusätzliches Kontextwissen zur Datenproduktion über andere Quellen verfügen.

Objektzentrierte Ablieferung

Bei der objektzentrierten Ablieferung verschiebt sich der Fokus von der Datenbank zu einzelnen Objekten, die aus einem Geschäftszusammenhang abgeliefert werden. In der funktionalen Architektur ist dieser Ansatz in der Applikationsschicht angesiedelt und unabhängig von der technischen Implementation der Datenbank in der Persistenzschicht. Beim objektzentrierten Ansatz werden entlang inhaltlich-geschäftslogischer Kriterien Ablieferungsobjekte gebildet. Diese Objekte werden entweder periodisch oder nach Abschluss einer spezifischen Geschäftsaktivität archiviert.

Die Ablieferungsobjekte können entlang unterschiedlicher Kriterien gebildet werden:

- Subjekte: Personen beispielsweise bei Einwohnerkontrolle, Studierendenverwaltung, Personalverwaltung
- Objekte: Daten zu Grundstücke, Gebäude, Fahrzeuge etc.
- Sitzungen: bei einer Sitzungsverwaltung mit Anträgen, Einladungen, Traktandenlisten
- Geschäfte: beispielsweise bei Einbürgerungen, Anträge
- Projekte: beispielsweise aus einer Fachapplikation für Projektportfolio-Management

¹⁰ Das zeigt sich auch an der Art, wie der Standard verwaltet wird. Die Versionen 1.0 und 2.0 wurden in der Schweiz als eCH-Standards publiziert, siehe <https://www.ech.ch/de/ech/ech-0165>. Die Folgeversionen wurden dann durch das DILCIS-Board publiziert:

<https://dilcis.eu/content-types/siard>.

¹¹ <https://www.bar.admin.ch/bar/de/home/archivierung/tools---hilfsmittel/siard-suite.html>

¹² <https://database-preservation.com/>

- Periode: beispielsweise bei Finanzbuchhaltung,
- Themen: beispielsweise dokumentarische, thematische Sammlung

In diesem Ablieferungsmodus werden Daten paketweise aus der Fachapplikation ins Archiv übertragen. Ob sie nach der Ablieferung aus der Fachapplikation gelöscht werden, muss von Fall zu Fall entschieden werden.¹³ Die erfolgte Ablieferung kann in der Applikation auch durch eine Statusänderung kenntlich gemacht werden. In jedem Fall muss verhindert werden, dass ein Löschvorgang in der Applikation zu Inkonsistenzen in der Datenbank führt.

Das Archiv kumuliert die übernommenen Objekte und setzt sie nach und nach zu einem umfassenden Bestand zusammen. Jedes Ablieferungsobjekt enthält sämtliche zum Objekt gehörenden Daten, Metadaten und Dateien in serialisierter Form. Die Objekte müssen in eine Form gebracht werden, welche die Speicherung und Nutzung unabhängig von der Ursprungsapplikation ermöglicht. Für die Speicherung im Archiv eignen sich dafür Formate, die auf semantischen Technologien und Linked Data basieren. Die Daten werden als Triples gespeichert und fügen sich nach und nach zu einem Knowledge Graph zusammen, der den abgelieferten archivwürdigen Teil der Fachapplikation repräsentiert.

Die objektzentrierte Ablieferung ermöglicht eine zeitnahe Archivierung einzelner Geschäfte unmittelbar nach deren Abschluss. Dies minimiert das Risiko nachträglicher Änderungen in der aktiven Datenbank. Innerhalb der Verwaltung kann das Archiv die archivierten Daten zum Zweck des Wissensmanagements bereitstellen. Sind die Daten öffentlich, können sie über geeignete Portale auch einer breiteren Öffentlichkeit zugänglich gemacht werden. Dieser Ansatz weist im Vergleich zur datenzentrierten Archivierung einer gesamten Datenbank auch den Vorteil auf, dass bei Benutzeranfragen nicht die gesamte Datenbank, sondern nur die nachgefragten Objekte aus dem Langzeitspeicher abgerufen werden müssen.

Applikationszentrierte Ablieferung

Beim applikationszentrierten Ansatz steht die Fachapplikation als Software-Artefakt im Fokus, wobei die obsoletere Fachapplikation als kulturelles Erbstück der Verwaltung verstanden wird. An ihr zeigt sich, wie Behörden Daten erfasst, bearbeitet, strukturiert und aufbewahrt haben. Die Mitarbeiter:innen einer Verwaltungsabteilung haben über Jahre hinweg eng mit dieser Applikation interagiert. Grafische Oberflächen, Tastatur-Shortcuts und Klickabläufe haben sich ins Gedächtnis eingebrannt und die Vor- und Nachteile der Applikation war Thema vieler Pausengespräche. Die Software ist ein Zeugnis, das aufzeigt, wie Verwaltungsarbeit in einer bestimmten Zeitperiode verrichtet wurde.

Für die applikationszentrierte Ablieferung stehen je nach Erhaltungsziel (*Preservation Intent*) unterschiedliche Ansätze zur Verfügung. Screenshots oder Screenshot-Videos helfen aufzuzeigen, wie die Mensch-Maschinen-Schnittstellen ausgestaltet waren. Bei eigenentwickelten

¹³ In öffentlichen Archiven macht je nachdem auch die Archivgesetzgebung Vorgaben, was wann gelöscht werden muss.

oder Open Source-Anwendungen kann der gesamte Programmcode archiviert werden, zusammen mit einer Anleitung zur Kompilierung für ein bestimmtes Betriebssystem und Prozessorarchitektur. Vorkompilierte Anwendungen werden als binäre Objekte archiviert, die mit Hilfe von Emulatoren wieder zum Leben erweckt werden können. Bei kommerziellen Anwendungen stellen sich hierbei Fragen des Urheberrechts, die mit der Lieferfirma geklärt werden müssen. Ein weit grösserer Aufwand stellt die Archivierung komplexer Systeme dar, die auf dem Zusammenspiel von Applikations- und Webservern, Datenbanken und Netzwerkkomponenten aufbauen, aber auch dafür gibt es Lösungsansätze (siehe Roeck et. al., 2019).

Angesichts dieses Fächers von Möglichkeiten erstaunt es, dass der applikationszentrierte Ansatz in der Diskussion um die Archivierung von Fachapplikationen jeweils relativ rasch verworfen wird. Begründet wird das sehr unterschiedlich, beispielsweise mit dem (falschen) Argument, dass die Daten aufgrund der Emulation nicht mehr nachnutzbar seien: «Mit der Emulationsstrategie könnte es möglich sein, das Programm selbst zusammen mit der Datenbank zu erhalten. Künftige Nutzende würden dann ein Angebot bekommen, das dem der heutigen Behördenmitarbeiterinnen und -mitarbeiter sehr nahekommt. Allerdings wären die Daten dann nicht mehr exportierbar und könnten somit nicht außerhalb dieser Emulationsumgebung verwendet werden. Dieser Ansatz wurde nach Kenntnis des Verfassers bislang noch von keinem Archiv umgesetzt.» (Keitel, 2017). Ein häufiges Argument sind auch die angeblich hohen Kosten die durch Softwarearchivierung entstehen. Zum Beispiel schreibt die Studiengruppe von nestor, die sich mit der Archivierung von Studierendenakten beschäftigt: «Es ist nicht vorgesehen, Funktionen des Fachverfahrens in einer informationstechnischen Anwendung lauffähig zu erhalten. Diese Entscheidung ist orientiert an dem Aufwand, der für den dauerhaften Systembetrieb in einer Emulation oder für die Erhaltung bestimmter Datenbankfunktionen erforderlich wäre. Die dazu erforderlichen Ressourcen sind für Hochschularchive weder gegenwärtig noch für die Zukunft absehbar» (nestor, 2022). Der Aufwand wird als hoch vermutet, aber tatsächlich kalkuliert wurde er offensichtlich nicht. Der Arbeitskreis "Archivische Bewertung" des Vereins deutscher Archivarinnen und Archivare spricht der Archivierung der Applikation gleich ganz den Sinn ab: «Die Bewertung elektronischer Fachverfahren konzentriert sich in erster Linie auf die in den Verfahren erfassten Informationen. Deren originale Umgebung kann (und muss in vielen Fällen auch) nicht auf Dauer erhalten werden.» (VdA-Arbeitskreis «Archivische Bewertung», 2014).

Die kategorischen Aussagen erstaunen, zumal aus den Textstellen nicht hervorgeht, dass vertiefte Untersuchungen zur Softwarearchivierung durchgeführt worden wären. Die Argumentation gegen die Archivwürdigkeit der Software stützt sich nicht auf inhaltliche Gründe, sondern auf den erwarteten hohen Aufwand. In anderen Kontexten, insbesondere bei Computerspielen und digitaler Kunst, vor allem Netzkunst, gilt Software durchaus als bewahrenswertes Kulturerbe. Im Bereich der Computerspielarchivierung hat sich eine engagierte Community etabliert, die grosse Anstrengungen unternimmt, um Emulatoren zu entwickeln, Original-Hardware zu erhalten oder sogar Hardware nachzubauen.

Computerspiele werden heutzutage als wesentlicher Bestandteil unseres kulturellen und audiovisuellen Erbes betrachtet. In der Schweiz finanzierte der Verein MemoriaV im Jahr 2022 ein Pilotprojekt, um die Situation der Erhaltung von Computerspielen auf nationaler Ebene zu untersuchen (Bernard et al., 2022). Das gegenwärtig laufende Forschungsprojekt "CH Ludens", das durch den Schweizerischen Nationalfonds gefördert wird, arbeitet die schweizerische Computergame-Industrie zwischen 1968 und 2000 auf.¹⁴ Im Bereich der Computerspiele wird die Nutzbarhaltung von historischem Programmcode als Grundanforderung an die Archivierung verstanden. Die Archivierungspraxis bei Fachanwendungen der öffentlichen Verwaltung weicht davon gegenwärtig deutlich ab.

Dies hat zur Folge, dass Fachanwendungen unbemerkt verschwinden, sobald sie den "End of Life"-Status erreichen. Heutzutage lässt sich kaum noch nachvollziehen, wie die textbasierten Benutzeroberflächen der 1980er Jahre oder die Fat Clients der 1990er Jahre ausgestaltet waren. Hinweise auf die Software finden sich in den Archiven höchstens in Unterlagen zu Beschaffungs- und Einführungsprojekten oder möglicherweise in Benutzerhandbüchern. Abgesehen davon muss eine Lücke in der Überlieferung festgestellt werden.

Zugang und Vermittlung

Wenn Daten aus Fachapplikationen im Archiv vorliegen, dann sollen sie auch genutzt werden, sei es durch die Verwaltung selbst im Rahmen des Wissensmanagements, sei es durch die breite Öffentlichkeit. Die Verwendung von XML- oder CSV-basierten Archivformaten, einschliesslich SIARD, bringt Herausforderungen mit sich, da diese Formate für die direkte Präsentation in einem Lesesaal oder einem Online-Portal wenig geeignet sind. Das nächste Kapitel beschreibt, wie archivierte Daten alternativ als Linked Data oder als Knowledge Graph verfügbar gemacht werden können. Hierfür werden die Archivdaten in ein RDF-basiertes Format konvertiert und über einen Triple Store zugänglich gemacht. Mit geeigneten Werkzeugen werden die Daten für die Nutzer:innen erfahrbar gemacht.

Linked Data als applikationsunabhängiges Nutzungsformat

Sollen Daten einer archivierten Fachapplikation im XML-basierten SIARD-Format genutzt werden, gibt es zwei Möglichkeiten: Entweder wird ein spezifischer Viewer bereitgestellt, oder die SIARD-Dateien werden zuerst in CSV- oder Excel-Tabellen konvertiert für die anschließende Auswertung. In beiden Fällen müssen die ganzen Datenbanken aus dem digitalen Langzeitarchiv geladen und für die Nutzung aufbereitet werden. Dieser Prozess ist besonders zeitintensiv bei

¹⁴ <https://chludens.ch/>

Datenbanken mit umfangreichen Datenmengen. Ohne Automatisierung ist dieser Vorgang im Archiv zudem mit erheblichem personellen Aufwand verbunden.

Als alternative Form bietet sich die Nutzung der archivierten Daten als Linked Data an, basierend auf dem «Resource Description Framework» (RDF). Hierfür werden die Daten in ein RDF-basiertes Format umgewandelt und in einen Triple Store geladen, eine auf Linked Data spezialisierte Datenbank. Die Nutzung erfolgt über eine standardisierte SPARQL-Schnittstelle oder Endbenutzer-Anwendungen, die auf diese Schnittstelle zugreifen.. Nach der Aufbereitung und Speicherung im Triple Store sind die Daten der archivierten Fachanwendung nutzbar. Abhängig von Schutzfristen und Datenschutzbestimmungen kann der Zugriff auf die Verwaltung beschränkt oder für die Öffentlichkeit freigegeben sein. Die Umwandlung von SIARD-XML zu Linked Data markiert einen wichtigen Schritt von der Archivierung im Sinne eines “Dark Archives” zur Archivierung als Erhaltung in einem anwendungsneutralen Format.

Im Bereich der Datenmodellierung bieten Linked Data gegenüber XML-basierten Formaten wie SIARD neue und leistungsfähigere Möglichkeiten. In RDF wird jede Ressource durch einen eindeutigen „Unique Resource Identifier“ (URI) identifiziert. Bei der Archivierung von Fachapplikationen bildet typischerweise ein Datenbank-Eintrag (Record) eine solche Ressource. Dank der URI und dem Subjekt-Prädikat-Objekt-Modell von RDF können die Datenbank-Records einer Fachapplikation in eine Netzwerkstruktur integriert werden, die weit komplexere Beziehungen erlaubt wie die Primärschlüssel-Fremdschlüssel-Beziehungen des relationalen Datenbankmodells oder die Baumstrukturen von XML. Diese Netzwerkstruktur erlaubt es, archivierte Daten mit Beschreibungen der zugrundeliegenden Geschäftsprozesse und Aufgaben sowie mit technischen Informationen zur Applikation zu verknüpfen. Es können auch Beziehungen hergestellt werden zwischen aggregierten Daten und Reports und den Rohdaten, auf denen sie basieren. Ein Archiv kann Daten unterschiedlicher Aggregationsstufen aus einer Fachapplikation entgegennehmen und diese untereinander verbinden. Entscheidet sich ein Archiv zusätzlich auch für eine applikationszentrierte Archivierung, können die Softwareartefakte nicht nur mit den Daten verknüpft werden, sondern auch mit den notwendigen Emulatoren und deren Konfigurationsdateien.

Das semantische Netz beschränkt sich auch nicht zwangsläufig auf die Binnenstruktur der einzelnen Fachapplikation. Es können Relationen zu Daten und Unterlagen definiert werden, die in Drittsystemen lagern, sowohl innerhalb wie auch ausserhalb der eigenen Organisation. Auf diese Weise lassen sich im Archiv auch komplexe Verwaltungs- und Systemstrukturen nachzeichnen, wie beispielsweise Geschäftsfälle oder Projekte, an denen mehrere Verwaltungseinheiten beteiligt sind, bei denen die Zuständigkeiten unklar oder aufgeteilt sind oder die Bearbeitung sich über mehrere Systeme hinweg erstreckt. Darüber hinaus können einzelne Datenbankeinträge mit externen Ressourcen und standardisierten Vokabularen wie der Gemeinsamen Normdatei (GND),¹⁵ dem

¹⁵ https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html

Virtual International Authority File (VIAF)¹⁶ oder, in der Schweiz, dem Historischen Lexikon¹⁷ verknüpft werden.

Publikation von SIARD als SIARD-RDF auf dem LINDAS-Portal

In der Schweiz verfügen sowohl der Bund als auch einige Kantone über die notwendige technische Infrastruktur, um Linked Data zu publizieren. Diese Infrastrukturen wurden aufgebaut, um “Open Government Data” in Form von Knowledge Graphen zu veröffentlichen. Beim Bund übernimmt das LINDAS-Portal (Linked Data Service) diese Aufgabe, dieses wird betrieben wird LINDAS durch das Schweizerische Bundesarchiv.¹⁸

Um archivierte Datenbanken über dieses Portal zu publizieren, müssen sie zunächst aus dem XML-basierten SIARD-Format nach RDF umgewandelt werden. Zu diesem Zweck haben das Schweizerische Bundesarchiv und die Fachhochschule Graubünden 2023 eine Ontologie für SIARD (SIARD-O) entwickelt und die Entitäten, Attribute und Relationen definiert, mit denen SIARD in RDF abgebildet werden kann. Dazu musste ein konzeptuelles Mapping von SIARD-XML nach SIARD-RDF definiert werden. Die Fachhochschule Graubünden entwickelte eine Konverter-Software, welche die Umwandlung komplett automatisiert vornimmt. Somit sind heute sowohl die Infrastruktur wie auch die Werkzeuge verfügbar für die Publikation archivierter Fachapplikationen in einem applikationsneutralen Format.

Das LINDAS-Portal stellt die Daten über eine standardisierte SPARQL-Schnittstelle zur Verfügung, die auch eine Volltextsuche in den Daten ermöglicht. Zudem verfügt das Portal über sogenannte Dereferenzierung. Wenn die URI einer Ressource über den Webbrowser aufgerufen wird, dann wird eine dynamische Website erzeugt, die eine textuelle Darstellung der Ressource und ihrer Attribute enthält. Die Attribute enthalten entweder Textwerte oder verweisen wiederum auf andere Ressourcen, die durch Anklicken der Links ebenfalls dereferenziert werden.¹⁹ Auf diese Weise kann die archivierte Fachapplikation Schritt für Schritt erforscht werden.

Wenn die archivierte Fachapplikation über einen SPARQL-Endpoint verfügbar sind, dann eröffnet dies Nutzungsmöglichkeiten, die weit über die Archivinstitution hinausreichen. Publierte Daten unterschiedlicher unterschiedlicher Archivinstitutionen können miteinander kombiniert und verbunden werden. Die Daten werden so Teil eines breiteren Ökosystems, wo sie mit Datensets unterschiedlicher öffentlicher und privater Provenienzen verbunden, aggregiert und neuen Nutzungszwecken zugeführt werden können.

Heute gilt das Provenienzprinzip nach wie vor als oberstes Kriterium archivischer Ordnung. Jedoch ist für die Nutzung der Einstieg über die Pertinenz (Personen, Ereignisse, Schlagworte, Geographie) oft viel hilfreicher und zielführender. Mit Linked Data können beide Zugänge gleichzeitig

¹⁶ <https://viaf.org/>

¹⁷ <https://hls-dhs-dss.ch/>

¹⁸ <https://lindas.admin.ch/>

¹⁹ <https://lindas.admin.ch/data-usage/data-usage-types>

bereitgestellt werden. Ein Weg dazu ist die Kombination der SIARD-RDF-Daten mit beschreibenden Metadaten im neuen Archivstandard "Records in Contexts" (RiC).²⁰ Denn anders als in den bisherigen Standards ISAD(G), ISAAR(CPF) und ISDF unterstützt RiC unterschiedliche Zugänge zum Archivgut (Wildi 2023).

Fazit

Der vorliegende Artikel ging davon aus, dass Fachapplikationen neben Records Management-Systemen den wichtigsten Quellenstrang in Archiven der öffentlichen Verwaltung bilden. Allerdings wurde festgestellt, dass es aktuell an Fachliteratur mangelt, die sich systematisch mit den Herausforderungen der langfristigen Nutzbarhaltung und Vermittlung von Daten aus Fachapplikationen beschäftigt. Dieser Beitrag zielt deshalb darauf ab, einige grundlegende Mechanismen der Archivierung von Fachapplikationen aufzuzeigen. Dabei liessen sich grundsätzlich drei Archivierungsstrategien unterscheiden: datenzentriert, objektzentriert und applikationszentriert. In der Praxis kann auch eine Kombination mehrerer Strategien sinnvoll sein. Für die Bereitstellung archivierter Fachapplikationen zeigt sich, dass semantische Technologien und Linked Data besonders geeignet sind. Ihr Hauptvorteil liegt darin, dass ein Zugriff auf einzelne Ressourcen, wie Datenbankeinträge, möglich ist, ohne die gesamte Datenbank aus dem digitalen Langzeitarchiv abrufen zu müssen. Nutzerinnen und Nutzer können über standardisierte Schnittstellen auf die Daten zugreifen und diese je nach Forschungsinteresse auch mit anderen Datenquellen kombinieren..

Literatur

Bernard, É., François, R., Krichane, S., Lüscher, A., Rochat, Y., Taddei, L., & Vetter, M. (2022). Pixelvetica—Sauvegarder le jeu vidéo suisse: État des lieux de la préservation du jeu vidéo en Suisse et dans le monde. <https://memoriav.ch/de/videogames-erhalten-pilotstudie-schlussbericht/>

Van Bussel, G.-J. (2017). The Theoretical Framework for the 'Archive-As-Is'. An Organization Oriented View on Archives. Archives in Liquid Times, Edition 1, Chapter II, Jaarboek 17. <https://www.researchgate.net/publication/321779123>

Ducry, E., Brodard, N., Broillet, L., Büchler, G., Guisolan, J., & Schneiter, P. (2018). Archivierung von Finanzdaten (ArFin).

²⁰ <https://www.ica.org/resource/records-in-contexts-conceptual-model/>

https://kost-ceco.ch/cms/dl/b544e3154f82a0ba2a2e1682a741e19b/17-036_ArFin_Schlussbericht.pdf

Hooft, V. (2023). Archiving by design: theory and practice in the Netherlands. In: Arbido 2023/1. <https://arbido.ch/de/ausgaben-artikel/2023/archiv-der-zukunft/archiving-by-design-theory-and-practice-in-the-netherlands>

Keitel, C. (2017). Fachverfahren (Südwestdeutsche Archivalienkunde). <https://www.leo-bw.de/themenmodul/sudwestdeutsche-archivalienkunde/archivaliengattungen/strukturierte-informationssammlungen/fachverfahren>

nestor-AG Archivstandards (Hg.) (2023). Nestor-Archivstandard. Archivierung von Studierendendaten aus Fachverfahren. Version 1.0. <https://d-nb.info/1294122746/34>

Niu, J. (2014). Appraisal and Selection for Digital Curation. International Journal of Digital Curation, 9, 65–82. <https://doi.org/10/gp5kpb>

Puchta, M., & Naumann, K. (Hrsg.). (2017). Kreative digitale Ablagen und die Archive: Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23.11.2016 in der Generaldirektion der Staatlichen Archive Bayerns (1. Auflage). Generaldirektion der staatlichen Archive Bayerns.

Roeck, C., Gieschke, R., Rechert, K., & Noordegraaf, J. (2019). Preservation strategies for an internet-based artwork yesterday today and tomorrow. <https://osf.io/2cqne/download>

Schweizerisches Bundesarchiv (2010). Gesamtkonzept für die Bewertung im Bundesarchiv. <https://www.bar.admin.ch/bar/de/home/service-publikationen/publikationen/archivfachliche-publikationen.html>

Schweizerisches Bundesarchiv (2013). Bewertungsempfehlungen. <https://www.bar.admin.ch/bar/de/home/service-publikationen/publikationen/archivfachliche-publikationen.html>

Schweizerisches Bundesarchiv (Hrsg.). (2022). SUPERB Bewertungsentscheid 2022. https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/gesamtkonzept_2010.pdf.download.pdf/gesamtkonzept_2010.pdf

VdA-Arbeitskreis «Archivische Bewertung». (2014). Bewertung elektronischer Fachverfahren. Diskussionspapier des VdA-Arbeitskreises „Archivische Bewertung“.

https://www.vda.archiv.net/fileadmin/user_upload/pdf/Arbeitskreise/Archivische_Bewertung/Bewertung_Fachverfahren_Positionen_StandDez2014.pdf

Wildi, Tobias (2023). Die Erweiterung des Provenienzprinzips: Der neue Records in Contexts-Standard. In: ARCHIV. theorie & praxis, 76. Jahrgang, Heft 03, Juli 2023, 166-173.

https://www.archive.nrw.de/sites/default/files/media/files/Archiv.theorie-praxisHeft3-2023Internet_0.pdf