

Protokoll zur 22. Tagung des VdW-Arbeitskreises
„Elektronische Archivierung“
am 24. November 2021 mit Teams

Thema

„Legen Sie’s doch einfach in die Cloud!“ Der Archivspeicher in einem OAIS-konformen Digitalen Archiv“

Nach einer Begrüßung durch Ulrike Gutzmann und einer Vorstellungsrunde der Beteiligten, bei der auch kurz auf die aktuell relevanten Fragestellungen hinsichtlich digitaler Archivierung bei den jeweiligen Archiven eingegangen wurde, folgte der Vortrag von Michael Kallookaran (fme).

TOP 1

Anforderungen an einen Archivspeicher

Michael Kallookaran ist von der IT in der Volkswagen AG beauftragt, ein Konzept zur Langzeitaufbewahrung zu erarbeiten. Das Archiv hat sich angeschlossen und den Auftrag um das Thema Langzeitarchivierung und Archivspeicher erweitert und dabei die Fragen gestellt: Ist das Konzept OAIS-konform, passt es für die digitale Langzeitarchivierung. Herr Kallookaran orientiert sich in seinem Konzept am OAIS-Modell, wie es im roten Magenta Buch von 2002 steht sowie an der deutschen Übersetzung in der Nestor-Veröffentlichung.

Die Präsentation (siehe Anhang) begann mit einer Übersicht zum OAIS-Modell, seiner Funktionsmodule und des Datenflusses. Im Fokus stand hier das Modul „Archivspeicher“ mit den Prozessen: Daten entgegennehmen, Speicherhierarchie verwalten, Fehlerkontrolle, Speichermedien ersetzen, Daten liefern, Notfallwiederherstellung.

Neben den Anforderungen des OAIS wurden die Anforderungen der IT an einen Archivspeicher präsentiert. Dabei handelt es sich um die IT-Schutzziele Vertrauenswürdigkeit, Integrität, Verfügbarkeit, Revisionsfähigkeit und Authentizität, die gewährleistet sein müssen.

Neben der Frage, was ein Archivspeicher zu leisten hat, ist auch zu definieren und genauer abzugrenzen, was ein Archivspeicher nicht ist und was somit durch andere Prozesse zu regeln ist. Ein Archivspeicher ist nicht für die Verständlichkeit der AIPs verantwortlich, sondern nur für deren Lesbarkeit. Die Verstehbarkeit der AIPs gilt es bereits einen Schritt vorher, d.h. beim Ingest, zu sichern. Die Dateiformate sind vor der Speicherung im Elektronischen Magazin festzulegen, auch im Hinblick auf das Modul Bestandserhaltung (Archivformate pdf/A, TIFF usw.). Wichtig ist, das Thema Ingest vorab zu klären bzw. immer im Zusammenhang mit dem Thema Archivspeicher zu sehen.

Bei Archivspeichern gibt es verschiedene Ausprägungen, die davon abhängen, wie teuer der Speicher ist, wie hoch die Verfügbarkeit sein muss, ob eine Datenbank

integriert ist oder nicht. Bei einem kostengünstigen Speicher sind die Zugriffszeiten auf die Daten länger. Aber hier ist das Bild des analogen Zwischenmagazins oder Außenlagers nützlich. Da dauert es häufig auch einen Tag, bis die analoge Akte geholt werden kann.

Wichtig ist bei der Anschaffung eines Archivspeichers, die Anforderungen des Archivs klar zu formulieren und die Begrifflichkeiten klar zu definieren. Wichtig ist auch, dass XML als Austauschformat möglich ist. Wichtig ist auch zu klären, wie man das Informationspaket (SIP) schnürt. Das muss beim Preingest und Ingest geklärt werden. Ein Archivspeicher alleine (Festplatten oder Tape) ist relativ „dumm“. Wichtig sind zusätzlich Viewer oder eine Datenbankverwaltung.

Zum Schluss seines Vortrags empfahl Michael Kallookaran die Auseinandersetzung mit dem E-ARK-Projekt, im Rahmen dessen auch hilfreiche Tools entwickelt wurden. Alle Archive, die noch am Anfang stehen, sollen auch bei den Open-Source-Produkten schauen.

Es entwickelte sich eine Diskussion, wer für das Thema Bestandserhaltung sowie die Datenkuratierung zuständig sei – IT oder Archiv. Diskutiert wurde, welche Möglichkeiten der Automatisierung bestehen und wie häufig eine Überprüfung erfolgen sollte. Dies kann in den einzelnen Archiven jeweils unterschiedlich definiert werden. Weiter wurde besprochen, was archiviert werden soll – Original und Langzeitformat oder nur Langzeitformat? Als grundsätzliche Regel wird definiert, dass sowohl das Original wie auch die Repräsentation archiviert werden.

Hervorgehoben wurde, dass begrifflich klar zwischen der „Langzeitarchivierung“ und der „Langzeitaufbewahrung“ unterschieden werden sollte. Während die Dokumente bei der „Langzeitarchivierung“ in Hinblick auf die Rechtssicherheit nicht mehr verändert werden dürfen – weshalb beispielsweise „word“ kein Langzeitarchivformat ist –, können Dokumente während der „Langzeitaufbewahrung“ mitunter noch verändert werden. Diese Trennung ist jedoch im Verständnis der einzelnen Abteilungen, die dem Archiv Dokumente abgeben möchten, nicht immer gegeben.

Eine nächste Frage, die diskutiert wurde, war, welche Daten übernommen werden können? Aus welchen Systemen? Mit welchen Metadaten? Ein Ziel wäre die Übernahme von Daten aus Records Management Systemen, die bereits strukturiert und mit Metadaten versehen sind. Auch stellte sich die Frage nach dem Archivspeicher selbst. Welche Erwartungen an Zugriffsgeschwindigkeit sind damit verbunden, wie langlebig muss er sein, welche Middleware ist erforderlich – was wiederum eine Frage der Kosten ist. Die Möglichkeiten und Budgets sind in den verschiedenen Archiven nicht einheitlich und bei Archiven ohne Anbindung an eine unterstützende IT stellt sich auch die Frage nach einer Cloudlösung als Speicherort, wenn kein eigener Archivserver gestellt werden kann. Innerhalb der Diskussion wurde auch dafür plädiert, dass gemeinsame Lösungen und Standards geschaffen werden, da Individuallösungen nicht für alle Archive im gleichen Maße finanzier- und stemmbar sind. Abschließend wurde festgehalten, dass die Kosten der Digitalen Archivierung nicht zu unterschätzen und noch nicht absehbar sind.

TOP 2

Wie geht's jetzt weiter? Fortsetzung der Arbeit im AKEA

Es stellte sich die Frage der Häufigkeit und Form der zukünftigen Sitzungen. Eine Fortsetzung von Teams-Sitzungen in kürzeren Abständen wurde begrüßt, was eine jährliche Präsenzsitzung aber nicht ausschließen sollte.

Weitere Diskussionspunkte und Anregungen:

- Definition der unterschiedlichen Rollen und Erwartungshaltungen der Archive
- Website-Archivierung wird ein zentrales Thema sein (welche Tools stehen zur Verfügung und sind bereits getestet worden, z.B. Pablo, MirrorWeb von Walter Nagel)
- Was können Archive bei der Digitalen Archivierung leisten und was nicht? (kreative Ablagen, riesige Mengen an Daten, Daten aus Systemen, Duplikate)
- Bewertung und Übernahme digitaler Fileablagen
- Archivierung von Daten aus Microsoft Teams

Aus der Diskussion ergaben sich konkrete Arbeitsgruppen, die sich selbst organisieren und ihre Ergebnisse 2022 präsentieren werden.

Rollen und Verständnis der Archive – Positionsbestimmung
Termin Ende Januar in Form einer digitalen „Kaffeerunde“
Koordination: Ulrike Gutzmann, Kornelia Rennert

AG 1 Webseitenarchivierung
Termin im Februar
Koordination: Daniel Wallburg

AG 2 Archivierung aus Microsoft Teams
Termin noch offen
Koordination: Normen Kremer

AG 3 Bewertung und Übernahme von digitalen Ablagen
Termin noch offen, ca. Ende April, Anfang Mai
Koordination: Bettina Hasselbring, Elke Pfnür

(Protokoll: Historisches Archiv des BR, 25.11.2021)