

Einführungsreferat Memoriav-Fachtagung 2019

Findmittel, Repository, Speicher: wie spielen diese Komponenten nach OAIS zusammen?

1 Abstract

Die Einführung zeigt die Entwicklung vom Katalog zum EDV-Katalog, vom Magazin zum Digitalen Magazin (Repository) und vom Speichermedium (Tape, Disk etc.) zum virtuellen Speicher und versucht, die Bezüge zur Begrifflichkeit von OAIS herzustellen.

2 Der Katalog

Die meisten Gedächtnisinstitutionen, d.h. Bibliotheken, Archive und Museen, werden ihren ersten Kontakt mit der Welt der Informatik bei der Einführung eines EDV-Katalogs als Ablösung vom Zettelkasten gemacht haben (Textverarbeitung für Inselprobleme gab es vielleicht schon früher). Aus dem EDV-Katalog wurde in den Archiven ein ganzes Archivinformationssystem, in den Bibliotheken ein umfassendes Bibliotheksverwaltungssystem.

Grundsätzlich hatte sich damit aber nichts Wesentliches geändert, verwaltet wurden weiterhin die Metadaten zu den Objekten der Gedächtnisinstitutionen.

Eine Vielzahl von Softwarelösungen hat sich hier etabliert, verschiedene Metadatenstandards haben sich entwickelt und durchgesetzt; Suchen, Finden, Verwalten bleibt aber der Hauptzweck.

Abgebildet wird dabei schon ein beträchtlicher Teil des OAIS-Modells, nämlich «Data Management» und «Access». Und weil die meisten Gedächtnisinstitutionen nun auch digitale Objekte in derselben Lösung verwalten möchten, haben etliche Hersteller ihre Kataloglösungen mit einem Datenspeicher verbunden, um *digitalen Ingest* und *Access* auf digitales Archivgut zu ermöglichen.

Nun bildet ein Katalog noch nicht vollständig das OAIS-Modell ab. Es fehlen (technische) Informationen zum Entstehungsprozess, zur Erhaltung und zur Gewährleistung der digitalen Objekte. Hier kommt das digitale *Repository*, «Archival Storage» nach OAIS ins Spiel.

3 Das Magazin

Das Magazin (oder engl. *Repository*) ist schon immer ein wichtiger Bestandteil der Gedächtnisinstitutionen, und es ist nichts neues, dass ein Magazin nicht nur eine ungeordnete Ablage ist, sondern gerade Archive verwalten einen nicht unerheblichen Metadatensatz direkt im Magazin (Ordnung, Beschriftung, Laufblätter etc.).

Bei der Entwicklung von «Digital Repository»-Lösungen hatten jedenfalls die Softwarehersteller eine vollumfängliche Implementierung des OAIS-Modells im Auge und keinesfalls eine Aufteilung der Metadatenverwaltung in Katalog und Magazin. Die meisten Produkte in diesem Bereich haben also auch eine Suche neben *Ingest*, *Storage*, *Access* und Verwaltungsfunktionen. Was hingegen (noch) nicht abgedeckt wird, ist die Verwaltung von physischen Objekten und allgemeine Managementaufgaben in Bibliotheken und Archiven.

Betrachten wir den softwaretechnischen Aufbau der verschiedenen Lösungen genauer, sehen wir, dass Katalogsysteme in der Regel auf einer Datenbank aufbauen und für die Interaktion ein oder mehrere Benutzerschnittstellen zur Verfügung stellen. Egal ob die Lösung aus einem installierten Programm (*Fat Client*)¹ und einer Datenbank oder eine *Multi-Tier* Lösung² (*Tier* engl. Schicht) mit Webinterface ist, werden die Metadaten nur in der Datenbank verwaltet.

Digitale *Repository*-Lösungen benötigen zwangsläufig Speicherplatz für die digitalen Objekte. Technisch wäre auch eine Speicherung der Objekte in einer Datenbank möglich, das hat sich aber bei den von uns betrachteten Lösungen nicht durchgesetzt. In der Regel werden zu den digitalen Objekten (z.B. eine TIFF-Datei) auch noch die vom OAIS-Modell verlangten Metadaten als Dateiobjekte auf diesem Speicher abgelegt. Für den Zugriff auf die Metadaten ist die Metadatenhaltung in Dateien eher ineffizient, darum werden bei allen betrachteten Systemen die Metadaten auch noch in einer Datenbank gehalten. Wir haben also eine Verdoppelung der Metadatenhaltung aus technischen Gründen (der Zugewinn an Sicherheit steht nicht im Vordergrund); die dateibasierten Metadaten sind aber im Gegensatz zu den Katalogsystemen die führenden (*leading*) Daten.

4 Der Speicher

So kommen wir zum Datenspeicher, der mithin wichtigsten Komponente des digitalen *Repository*. Der Speicher wird als Dateispeicher von einem Speichersystem zur Verfügung gestellt, heute als Dateisystem mit Ordnern und Dateien, beides in der Regel in Grösse und Anzahl eingeschränkt. Weil nun diese Dateien oder Bits doch auf einem physischen Medium magnetisch oder elektronisch gespeichert werden müssen und jeder physische Träger (Tape, Disk, etc.) endlich ist, virtualisiert das Speichersystem den begrenzten Speicherplatz eines Mediums und fügt viele Medien zu einem grossen Datenspeicher zusammen. Aus Gründen der Sicherheit befinden sich die gleichen Daten in so einem Speichersystem in der Regel nicht nur auf einem Medium, sondern verteilen sich über viele Medien und möglicherweise auch Rechenzentrum-Standorte. Bei hochentwickelten Speichersystemen kann auch erwartet werden, dass die Medien am Ende ihrer Lebenszeit (heute bei Disk und Tape etwa 5 Jahre) ausgetauscht werden können, ohne das sich bei der logischen/applikatorischen Sicht auf den Speicherplatz etwas verändert³.

Wie wir jetzt verstehen, ist es nun vom Speicherplatz, der von einem lokalen Speichersystem verwaltet wird, zum Cloud Speicher ein kleiner Schritt. Der Cloud Speicher ist einfach etwas weiter weg, aber genauso wie das lokale System (wir kennen es als File-Server) via Netzwerk erreichbar. Ein Cloud Speicher öffnet aber auch die Möglichkeit, dass der Speicherbedarf vieler Nutzer gebündelt werden kann und im Sinne der «Economy of Scale» für die einzelnen günstiger wird.

¹ Ein *Fat Client* ist ein Programm, das auf jedem Benutzer PC installiert ist und seine Daten von einem zentralen Datenbankserver bezieht (die ältere, klassische, Applikationsarchitektur von sog. Fachapplikationen).

² *Multi-Tier*-Lösung bestehen aus einem Datenbankserver, einem Applikationsserver und einem Webserver. Der Benutzer interagiert mit dem Browser mit dem Webserver, dieser holt sich die eigentliche Applikation vom Applikationsserver, der Applikationsserver wiederum verwaltet die Daten auf dem Datenbankserver (Beispiel www.sbb.ch)

³ Also keine Migration von einem Datenspeicher auf einen anderen, neuen Datenspeicher notwendig ist.

5 Die Prozesse

Die eigentlichen Prozesse, wie sie im OAIS Referenzmodell⁴ beschreiben sind, und ihr Zusammengehen mit den heutigen technischen Umsetzungen sind jetzt noch nicht detailliert beschrieben.

Die Ablieferung/*Submission* digitaler Daten erfolgt in einem *Submission Information Package* (SIP), im SIP sind die digitalen Daten und Metadaten enthalten. Im nachfolgenden *Ingest* Prozess werden die abgelieferten Daten und Metadaten geprüft und mit weiteren, prozessualen und technischen Metadaten ergänzt.

SIP und diese zusätzlich im Ingest generierten Metadaten werden im *Archival Information Package* (AIP) gespeichert. Das AIP wird anschliessend im Repository, dem *Archival Storage* abgelegt.

Im Ingest wird bei den heutigen Systemlösungen in der Regel automatisch oder halbautomatisch ein Metadatensatz (*Descriptive Information*) an den bereits vorhandenen Katalog übergeben.

Eine Suche im Katalog nach digitalem Archivgut führt zum Zugriff auf ein AIP, das AIP wird für den Benutzer zum speziellen Informationsobjekt, dem *Dissemination Information Package* (DIP) überführt, das je nach Zugriffsrechten und Benutzungszweck zum AIP unterschiedlich aufgebaut sein kann.

⁴ Hier die gemeinhin gezeigte Grafik, daneben die prozessbezogene Darstellung <https://public.ccsds.org/pubs/650x0m2.pdf> Seite 4-1 und 2-8

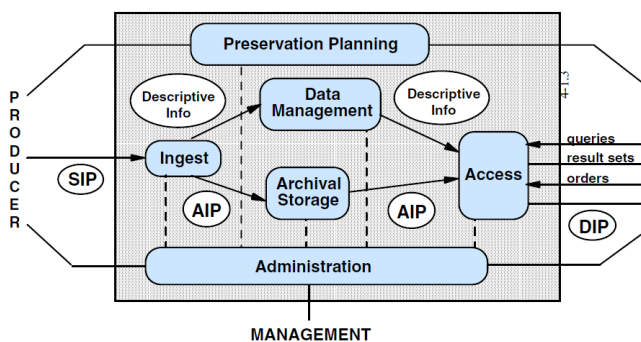


Figure 4-1: OAIS Functional Entities

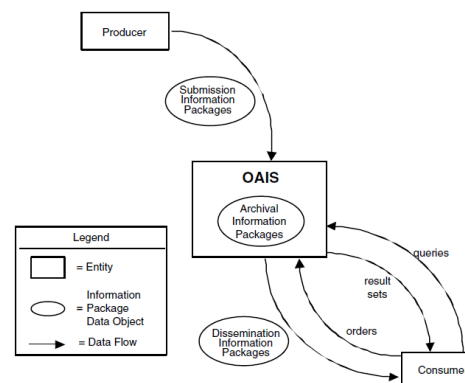


Figure 2-4: OAIS Archive External Data