

Normen und Standards als Synergiepotentiale in der digitalen Archivierung

Dr. Tobias Wildi, Geschäftsführer Docuteam. Kontakt: t.wildi@docuteam.ch

Publiziert in Arbido 2/2012. Bitte nach der gedruckten Version zitieren.

Abstract

Ein zentrales Element des Open Archival Information System (OAIS) ist das Datenmodell, das darauf beruht, dass nicht einzelne Dateien, sondern logische Informationsobjekte mit Daten und Metadaten langfristig nutzbar gehalten werden. Für die technische Umsetzung des Informationsmodells existieren heute zahlreiche Standards und Formate, die aufeinander abgestimmt werden müssen. Docuteam schlägt eine Architektur vor, die auf bestehenden Standards aufbaut und diese miteinander kombiniert. Die Nutzung bestehender Standards und Formate macht auch die Nutzung bereits bestehender Open Source-Software möglich. Dies senkt die Kosten in der Umsetzung eines OAIS und macht digitale Archivierung auch für kleine Institutionen möglich.

Der deutsche Archivar Nils Brübach schrieb 2002 in einem Tagungsbeitrag über das damals gerade neu publizierte Modell des Open Archival Information System (OAIS): «Die kleinen, auf PC-Basis realisierten Einzellösungen werden in der Zukunft nicht mehr möglich sein, weil sie den zentralen Punkt der Nachhaltigkeit der zu entwickelnden Lösung und ihrer Vernetzbarkeit nicht gewährleisten.»¹ Seit der Publikation dieses Papiers sind zehn Jahre vergangen und in dieser Zeit hat sich das OAIS zur allgemein anerkannten Referenz entwickelt. Rund um das Modell hat sich eine ganze Landschaft von Normen, Standards und Formaten entwickelt, die einzelne Teile des OAIS konkretisieren und im Detail spezifizieren. Der vorliegende Artikel lotet aus, wie die einzelnen Archivinstitutionen von diesen Standardisierungsbemühungen konkret profitieren können und prüft, ob und wie in der Zwischenzeit das «OAIS für kleine Einheiten» doch noch möglich wurde.

Standardisierung der Informationsobjekte

Ein Zentrales Element des OAIS ist sein Datenmodell. Dieses basiert auf der Logik von Informationsobjekten, die langfristig nutzbar gehalten werden. Unterschieden werden drei Aggregatzustände zum Zweck der Datenübergabe (*Submission Information Package*), der Archivierung (*Archival Information Package*) und der Vermittlung (*Dissemination Information Package*). Ein Objekt setzt sich, wie Abbildung 1 zeigt, aus den hier lediglich im Telegrammstil genannten Elementen zusammen: Nutzdaten, die archiviert werden sollen (*Data Object*). Eine Anleitung, wie diese Daten zu einem späteren Zeitpunkt interpretiert und genutzt werden können (*Representation Information*). *Preservation Description Information* dokumentieren die

¹ Brübach, Nils. OAIS – Das «Open Archival Information System»: Ein Referenzmodell zur Organisation und Abwicklung der Archivierung digitaler Unterlagen. Publiziert in den Tagungsunterlagen des Arbeitskreis «AudS», 6. Tagung, 5.–6. März 2002, Dresden. Publiziert unter: <http://www.staatsarchiv.sg.ch/home/auds/06>

Herkunft und Integrität des Informationsobjekts, sowie die Massnahmen des Archivs zum Zweck der langfristigen Nutzbarhaltung. Die *Packaging Information* (Container) setzt Daten und Metadaten zueinander logisch oder physisch in Verbindung setzt. Zudem braucht es weitere Metadaten für die Inhaltsbeschreibung (*Descriptive Information*).

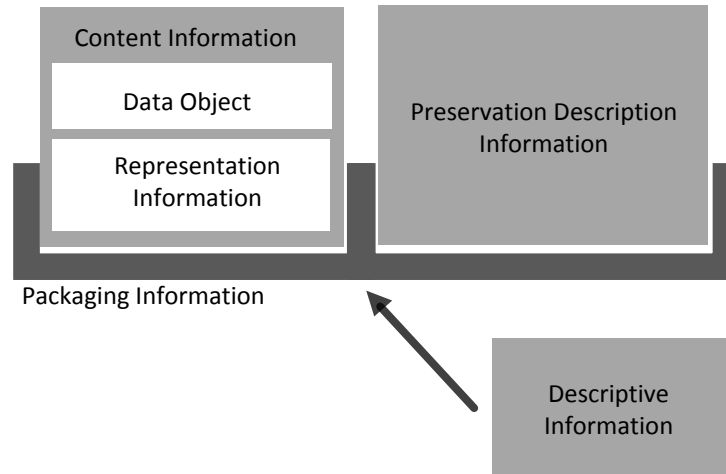


Abbildung 1: Informationsmodell des OAIS

Das OAIS gibt dieses Informationsmodell zwar vor, setzt aber keine Leitlinien bezüglich technischer Umsetzung. Entsprechend entstand das Bedürfnis nach detaillierteren Standards für die einzelnen Elemente des Datenmodells, bis hin zu technisch umsetzbaren Formaten. Für die Nutzdaten entstanden Kataloge mit Dateiformaten, die für die Archivierung besonders geeignet sind, in der Schweiz etwa beim Bundesarchiv (BAR) oder bei der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST). *Representation Information* wird für zahlreiche Dateiformate in zentralen Registraturen hinterlegt, prominentestes Beispiel ist PRONOM des britischen Nationalarchivs. Standards wie PREMIS oder LMER bilden *Preservation Description Information* ab. Schliesslich werden Formate wie METS oder XFDU für Objektcontainer verwendet. Für beschreibende Metadaten existieren neben den Standards des Internationalen Archivrats (ISAD, ISAAR, ISDF und ISDIAH) die kanadischen *Rules for Archival Description* (RAD) und der amerikanische DAC-Standard. Für alle diese Standards dienen EAD und EAC als unabhängige Austauschformate.

Die Aufzählung zeigt, dass zwar für jeden Einzelaspekt des Informationsmodells zahlreiche Standards und Formate existieren, diese sind aber nicht aufeinander abgestimmt, respektive überschneiden sich in Teilbereichen sogar. Beim Aufbau eines OAIS kann eine Institution nicht einfach einen einzigen Standard «Out of the Box» verwenden, sondern muss sich aus dem oben genannten Angebot die Struktur seines Informationsobjekts selber zusammenstellen. Dies ist eine herausforderungsreiche Tätigkeit und es soll an zwei Beispielen aufgezeigt werden, wie dies in der Schweiz gelöst wurde.

Beim ersten Beispiel wählte das BAR und in der Folge auch die KOST in ihrer SIP-Definition den radikalen Ansatz, dass sie sich bei den Nutzdatenformaten zwar auf weit verbreitete Standards stützten, für die Metadaten jedoch ein eigenes proprietäres Format definierten. Dessen Struktur ist passgenau auf die momentanen Bedürfnisse der Bundesverwaltung zugeschnitten, das heisst auf die Archivierung von Unterlagen primär aus Geschäftsverwaltungssystemen. Der Vorteil ist, dass die Struktur zumindest gegenwärtig noch relativ einfach gehalten werden kann. Momentan hat das BAR-SIP noch Defizite, indem wichtige Teile des OAIS-Informationsmodells nicht abgebildet werden können. Beispielsweise

fehlt die Möglichkeit zur Hinterlegung sowohl von *Representation Information* als auch weitgehend von *Preservation Description Information*. Beispielsweise können keine Workflowmetadaten aus einem Records Management-System mitarchiviert werden, was den Authentizitätsnachweis digitaler Ablieferungen langfristig sehr schwierig macht. Beim zweiten Beispiel hat Docuteam bei der Definition eines Objektmodells den umgekehrten Weg gewählt, indem die Prämisse gewählt wurde, «nichts Neues zu erfinden». Im Gegensatz zu einem Nationalarchiv besitzt das Kleinunternehmen keinerlei normgebende Kraft. Docuteam ist von bereits bestehenden Standards und Formaten ausgegangen und erarbeitete eine Architektur, die das OAIS-Informationsmodell wie folgt abbildet: Die Nutzdatenformate sind offen und je nach Anwendungsfall zu wählen, für *Representation Information* wird auf die Fileregistry des britischen Nationalarchivs (PRONOM) referenziert, der Premis Data Dictionary findet für alle *Preservation Description Information* Anwendung, METS bildet den Container und zur Beschreibung wird EAD (Encoded Archival Description) verwendet. Auf den ersten Blick scheint es sich um eine eher schwierig zu orchestrierende Zusammenstellung zu handeln – gerade wenn man das viel einfacher strukturierte Beispiel des BAR daneben stellt. Bei näherer Betrachtung ergeben sich aber zwei grosse Vorteile. Das Objektmodell von Docuteam beschränkt sich nicht auf den Anwendungsfall der Geschäftsverwaltung, sondern ist für beliebige Objekttypen einsetzbar. Es werden damit bereits heute neben Verwaltungsakten auch Bibliotheksbestände, digitale Kunstwerke, ganze Websites oder audiovisuelle Kulturgüter archiviert. Weil «nichts Neues erfunden», sondern international bereits weit verbreitete Standards und Formate genutzt und rekombiniert wurden, reduzieren sich als Nebeneffekt auch die Kosten zur technischen Umsetzung. Auf diesen besonderen Aspekt geht der nächste Abschnitt ein.

Entwicklung von Werkzeugen

Wenn sich in mehreren Archiven und Bibliotheken gleichzeitig bestimmte Standards und Formate durchsetzen, dann eröffnet sich auch die Möglichkeit zur Wiederverwendung und Standardisierung von Werkzeugen. Dass dies die Kosten zum Aufbau und Betrieb eines digitalen Archivs beträchtlich reduziert, ist eine banale Feststellung. Sie wird aber bei der Diskussion um die Vor- und Nachteile bestimmter Standards kaum je angeführt und soll hier als besondere Facette hervorgehoben werden.

Die These «Standardisierung = Kosteneinsparung» kann beispielhaft am METS-Standard der Library of Congress aufgezeigt werden. Es ist nicht von der Hand zu weisen, dass METS als Standard durchaus Defizite aufweist. METS ist komplex strukturiert und entsprechend aufwändig in der technischen Umsetzung. Die Trennung zwischen Containerinformationen und anderen Bereichen des OAIS-Informationsmodells ist zu wenig scharf gezogen und zudem fehlen bestimmte Verlinkungsmöglichkeiten zwischen den einzelnen Abschnitten. Und trotzdem hat sich METS in den letzten Jahren weltweit in zahlreichen Bibliotheken und Archiven durchgesetzt. Entsprechend sind auch Werkzeuge und Programmbibliotheken zur Verarbeitung METS-basierter Informationsobjekte entstanden. Da sind mächtige kommerzielle Produkte wie Rosetta von Ex Libris oder die «Safety Deposit Box» von Tessella zu nennen. Da blüht aber auch eine farbigere Wiese von Open Source-Initiativen zur Erzeugung, Editierung, Verwaltung und Präsentation von METS-Objekten.² Der frei verfügbare Programmcode ermöglicht es heute auch kleinen Archivinstitutionen, ein eigenes OAIS aufzubauen, zu betreiben und laufend für neue Aufgaben zu erweitern. Beispielsweise hat das Stadtarchiv Baden bereits 2008 als erstes schweizerisches Kommunalarchiv mit dem Aufbau eines OAIS begonnen. Basierend auf dem

² Eine Übersicht bietet <http://www.loc.gov/standards/mets/mets-tools.html>

oben beschriebenen Docuteam-Objektmodell und unter Verwendung von Open Source-Werkzeugen konnte dieses Projekt mit einem beschränkten Budget angegangen werden. Dass dieses Vorgehen nicht nur für kleine Archive funktioniert, sondern durchaus skalierbar ist, zeigen die nach dem gleichen Prinzip aufgebauten digitalen Archive in den Staatsarchiven Wallis und Basel Landschaft, dem Stadtarchiv St. Gallen und den Unternehmensarchiven von ABB Schweiz und SRG SSR.

Welche Werkzeuge werden nun aber konkret eingesetzt? Drei Open Source-Werkzeuge sollen hier namentlich genannt werden. Abbildung 2 zeigt Docupack, ein von Docuteam entwickeltes Werkzeug zum Anlegen, Editieren und Visualisieren von Informationsobjekten.³ Die Software basiert auf den oben beschriebenen Standards METS, PREMIS und EAD.

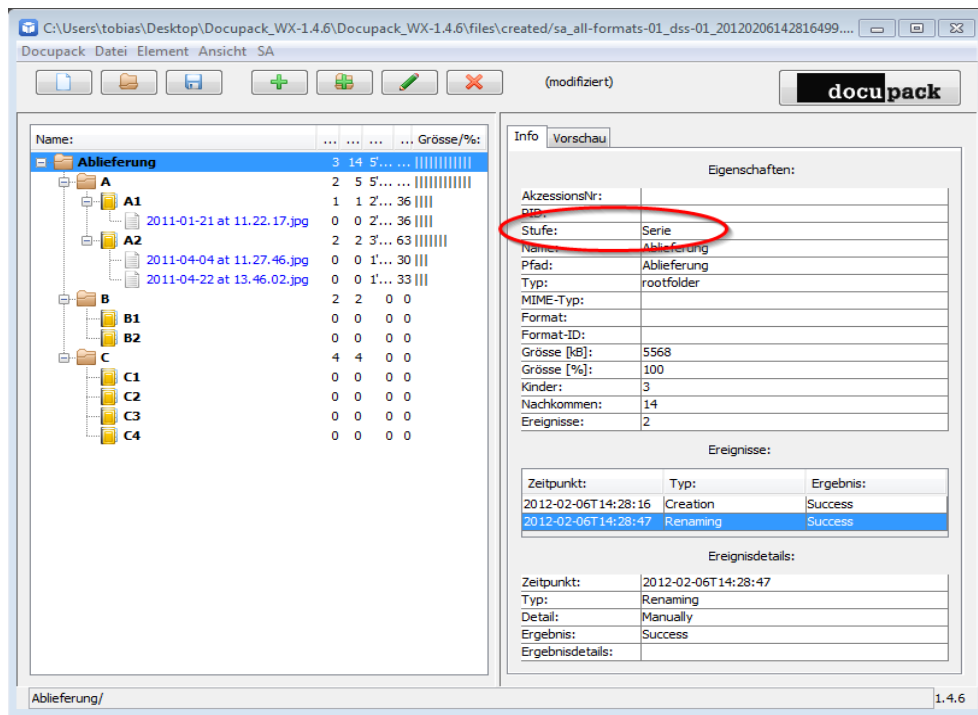


Abbildung 2: Docupack als Editor für Informationsobjekte.

Abbildung 3 ist ein Screenshot von ICA AtoM, welches die ICA-Standards für beschreibende Metadaten abbildet und im Sinne einer Referenzanwendung aufzeigt, wie diese miteinander verbunden sind.⁴ In der Westschweiz erfährt das Werkzeug gegenwärtig eine starke Verbreitung, mangels Übersetzung der Benutzeroberfläche sind keine Installationen in der Deutschschweiz bekannt.

³ <http://wiki.docuteam.ch>

⁴ <http://www.ica-atom.org>

ICA AtoM Demo site

Advanced search

[archival descriptions](#) [authority records](#) [archival institutions](#) [functions](#) [subjects](#) [places](#) [digital objects](#)

Add ▼
Manage ▼
Import ▼
Admin ▼

View archival description

Fonds 98435 - Maurice Lamontagne fonds
📄

Identity area

Reference code	CA LAC 98435
Title	Maurice Lamontagne fonds
Date(s)	○ 1943-1-1 - 1982-1-1 (Creation) Lamontagne, Maurice
Level of description	Fonds
Extent and medium	2.3 m of textual records 1 drawing: gouache on pencil

Context area

Name of creator	Lamontagne, Maurice (1917-1983)
Biographical history	
Maurice Lamontagne was born in Montréal. He conducted his studies at the Rimouski Seminary and the University	

Archival institution

Library and Archives Canada

Creator(s)

○ [Lamontagne, Maurice](#)

Fonds

- [Fonds 98435 - Maurice Lamontagne fonds](#)

+ [Series 109431 - Speeches](#)

+ [Series 109432 - Articles and memoranda](#)

Export

○ [Dublin Core 1.1 XML](#)

○ [EAD 2002 XML](#)

Abbildung 3: ICA AtoM zur Verwaltung beschreibender Metadaten

Und weiter ist Fedora Commons zu nennen, ein Repository-System für Bibliotheken und Archive mit einer sehr flexiblen Objektarchitektur. Fedora eignet sich bestens, um METS-basierte Objekte entgegenzunehmen, zu speichern, verwalten und zugänglich zu machen.⁵ Zusammen decken diese drei Werkzeuge die Grundanforderungen eines OAIS weitgehend ab.

Das «OAIS für kleine Einheiten»

Um auf das Ausgangszitat von Nils Brübach zurückzukommen: Die These, dass sich ein «Ein-PC-OAIS» nicht realisieren lässt, ist längst nicht mehr aufrecht zu halten – übrigens hat Brübach selber diese Aussage in späteren Beiträgen auch abgeschwächt. Heute ist es auch für Organisationen mit bescheidenen Ressourcen und IT-Infrastrukturen möglich, ein eigenes digitales Archiv aufzubauen und zu betreiben – wenn nötig auch auf einer einzigen physischen Maschine. Ausgangspunkt dafür ist, sich auf Standards und Formate abzustützen, die international weit verbreitet und anerkannt sind. Damit eröffnet sich einerseits der Zugang zu einer breiten Palette von Werkzeugen, aber auch die Möglichkeit zum direkten Erfahrungsaustausch mit Fachkollegen, die ihre digitalen Archive auf ähnliche Weise betreiben. Diese Synergiepotentiale gilt es aktiv zu nutzen.

⁵ <http://fedora-commons.org/>