

Auszug aus:

„Bestandsanalyse der Datenbanken des Verbund Deutsche Tanzarchive und Konzept zur Institutsübergreifenden Vernetzbarkeit“ (April 2012)

Dipl. Inf. Michael Christen

1 Einleitung

Der Verbund Deutsche Tanzarchive („VDT“) plant im Zusammenschluss für die nächsten Jahre ein Projekt zur Erhaltung, Erschließung, Digitalisierung und Vernetzung ihrer vielfältigen Archivalien aus dem Print-, audiovisuellen und aus anderen Bereichen der vielfältigen Zeugnisse dieser Kunstsparte. Diese Expertise soll sowohl eine qualitative (im Hinblick auf Datenmodellierung) wie auch eine quantitative Bewertung der gegenwärtigen (jeweils höchst unterschiedlichen) Datenbanken möglich machen und Arbeitsschritte für die Erreichung einer Lösung aufzeigen. Dazu werden Erfahrungen im Umgang mit Datenbanken zum Thema Tanz sowie Anforderungen zur Visualisierung und Kategorisierung der Daten dieser Datenbanken im Verbund der Tanzarchive reflektiert. Der Verbund Deutsche Tanzarchive möchte diese Expertise als Entscheidungsgrundlage für die nächsten Schritte bei der Vernetzung der Archive und deren wünschenswerte gemeinsame Internetpräsenz nutzen und als Vorbereitung der daraus folgenden Workshops. Daher wird in diesem Dokument auf eine Auflistung von Arbeitsschritten zur Erreichung der genannten Ziele hingearbeitet.

1.1 Herausforderungen

Die Datenbanken der Archive des VDT sind historisch gewachsen und bauen auf ganz unterschiedlichen institutionellen Kontexten auf. Die Arbeit der Archive könnte durch eine Konsolidierung der Daten und des darin enthaltenen Wissen erheblich profitieren, jedoch ist einerseits eine einfache Zusammenlegung der Datenbanken aufgrund der ganz unterschiedlichen institutionellen Aufgaben ausgeschlossen und andererseits die Zusammenführung der Daten aufgrund der großen Vielfalt und Unterschiedlichkeit der in den Archiven verwalteten Objekte mit bislang bekannten Methoden nicht zu bewältigen. Die Archivare des VDT wünschen sich eine Vernetzung der vielfältigen Objekte ohne die bereits erfasste Detailgenauigkeit der Objektbeschreibungen zu verlieren. Diese höchst sinnvolle, aber anspruchsvolle Anforderung an eine Vernetzung der Daten der Archive kann mit Hilfe von Wissensmanagement-Methoden, Verfahren im Umgang heterogener Daten und dem Nutzen von spezialisierter Software ermöglicht werden. Die genannten Anforderungen verfeinern sich zu folgenden Gesichtspunkten:

- Es wird eine Beschreibung benötigt, ob und wie eine tragfähige Datenstruktur für eine konsolidierte Datenbasis der Tanzarchive hergestellt werden kann. Aufgrund von Vorgehensweisen im Wissensmanagement ist hierfür eine Ontologiedefinition für das Themengebiet Tanz nützlich. Eine

solche Ontologie für das Themengebiet Tanz existiert aber bislang nicht und muss als akademische Aufgabe erarbeitet werden. Die Schaffung einer Ontologie könnte das Dach bilden, unter dem die Eigenständigkeit der Archive erhalten bleibt bei gleichzeitig fruchtbarer Vernetzung und gemeinsamer Außenpräsenz.

- Bei der Errichtung der Deutschen Digitalen Bibliothek mussten Daten aus den Bereichen der Bibliothek, der Archive und der Museen konsolidiert werden. Es hat sich dabei gezeigt, dass die zu vernetzenden Institutionen bestimmte Vorarbeiten durchführen mussten, damit aufgrund der Verschiedenartigkeit der Objektarten die Metadaten der Objekte konsolidiert werden konnten. Beispielsweise haben das Institut für Museumsforschung Berlin, das Bildarchiv Foto Marburg und das Zuse-Institut Berlin für den Bereich Museum ein eigenständiges Vokabular ‚museumdat‘ erstellt (<http://www.museumdat.org/>), um eine Vernetzungsfähigkeit untereinander und mit Bibliotheken und Archiven herzustellen. Die Vorgehensweise bei diesem Vorhaben soll ein Vorbild für die Konsolidierung der Tanzarchive sein und bei der Harmonisierung der Daten zur Anwendung kommen. Aufgrund der Verschiedenartigkeit der Objektarten in den Tanzarchiven des VDT müssen die Archive des VDT gemäß dem Vorbild ‚museumdat‘ ebenfalls eine Vokabular- und Termdefinition vornehmen, damit eine Klassifikation der Tanzarchiv-Objektarten innerhalb einer vernetzten Plattform möglich wird.
- Es soll die Möglichkeit betrachtet werden, das Ergebnis der Harmonisierung der Tanzarchiv-Daten als Quelle für die „Deutsche Digitalen Bibliothek“ („DDB“) zu nutzen. Die Möglichkeit, die Daten der Tanzarchive in die DDB zu integrieren, würde gleichzeitig bedeuten, dass die Software der DDB für ein (Such-) Portal der Tanzarchive eingesetzt werden könnte. Es wird daher angestrebt, die Vernetzung der Tanzarchive u.a. mit Hilfe der Software ‚Cortex‘ der DDB zu realisieren.
- Um fachlich korrekt vorzugehen müssen die beteiligten Archive ihre Daten schrittweise in ein angleichbares Format bringen, ohne ihre eigene Selbstständigkeit und Aktionsfähigkeit damit einschränken zu müssen. Die notwendigen Maßnahmen zur Erreichung des Zieles sollen so weit im Detail skizziert werden, dass es jedem Archiv des VDT im Rahmen des Projekts möglich wird, diese nächsten Schritte zu planen. Das Ergebnis dieser Expertise soll sein, die notwendigen Schritte so weit detailliert zu erläutern, dass ein darauf aufbauender Projektantrag möglich wird.

2. Arbeitsschritte

Arbeitspunkt (1) Abgleich des gemeinsamen Tanz-Vokabulars

Zu Beginn der Arbeiten sollte eine Plattform zur Koordination der Vokabularienbildung gefunden werden. Dies wurde im Kontext der Erstellung dieser Expertise bereits exemplarisch durch die Bereitstellung eines Wikis als gemeinsame ‚Werkbank‘ für die Mitarbeiter der Archive angeregt.

Arbeitspunkt (2) Datenexport aus den bisherigen Datenbanken des VDT

Für jedes bisherige Datenbanksystem muss eine geeignete Exportmethodik gefunden werden.

Arbeitspunkt (3a) und (3b): fachliche Darstellung und Transformation

Es ist eine fachliche Darstellung

- (3a) für die Wissensdaten und
- (3b) für die Objektdaten zu finden

Die fachliche Darstellung wird fortan das ‚Zielformat‘ genannt. Die Datenbanken können nicht das gewünschte Zielformat direkt beschreiben. Es ist jedoch möglich, die Exporte aus Arbeitspunkt (2) zu dem Zielformat mit Hilfe einer Software zu transformieren. Das Zielformat soll alle Ursprungsdaten vollständig wiedergeben.

Für die fachliche Darstellung des Wissens und der Objektdaten kann in dieser Expertise noch keine genaue Definition gegeben werden. Eine Empfehlung für ein fachlich richtiges Datenformat entsteht aus der akademischen Auseinandersetzung mit Tanzthemen und deren Formulierung in einer Ontologie. Dies sollte als nächster Schritt durchgeführt werden.

Arbeitspunkt (3a) Empfehlung für die Ablage von Wissensdaten

Die transformierten Wissensdaten liegen nach der Transformation in einem RDF Schema vor. Die in diesem Format vorliegenden Daten sollen dann so abgelegt werden, dass sie dort bearbeitet werden können. Damit diese Daten leicht zu bearbeiten sind, wäre eine Eingabemaske, ähnlich einer Datenbankmaske, geeignet. Die Frage, wie diese gestaltet werden kann und mit welcher Software, ist weniger wichtig als die Tatsache, dass die Daten weiterhin als RDF verfügbar sind, damit sie in die Suchmaschine eingespielt werden können. Eine möglicherweise einfach zu errichtende Verwaltungsform könnte mit einem semantic Wiki, beispielsweise mit dem semantic media wiki, bereitgestellt werden. Die Errichtung einer Eingabemöglichkeit gehört zu den nächsten Schritten die unmittelbar auf die Definition einer Wissens-Ontologie für den Themenbereich Tanz folgt.

Arbeitspunkt (3b) Empfehlung für die Ablage von Objektdaten

Die transformierten Objektdaten liegen nach der Transformation ebenfalls in einem RDF Schema in Form von einzelnen Dateien vor (pro Datenbankeintrag eines). Die heutigen Filesysteme arbeiten intern ähnlich wie eine Datenbank und die Ablage von mehreren Millionen Dateien ist unproblematisch.

Arbeitspunkt (4) Entwicklung einer Ontologie für die fachliche Darstellung des Wissens und der Objekt-Daten

Dieser Vorgang sollte gleichzeitig zu (3) der Bestimmung eines fachlichen Datenformats gestartet und vorgenommen werden. Dieser Arbeitspunkt hat die Vereinheitlichung von einzelnen Datenattributen der Datenschemata zum Ziel: es müssen solche Datenattribute identifiziert werden, welche miteinander gemappt werden können. Hier ist auch festzustellen, ob bestimmte Datensätze Identifikatoren besitzen, die auf andere Objektattribute vererbt werden können. Als Ergebnis dieses Arbeitspunkt muss erarbeitet werden:

- eine Transformationsregel zur Übersetzung des fachlichen Datenformats für das Wissen und die Objektdaten nach CIDOC-CRM
- eine ‚Transitions‘-Regel zum Mapping der Daten im CIDOC-CRM Modell.

Arbeitspunkt (5) Export der Wissensdaten aus dem Dateneditor nach RDF

Daten, welche in semantic media wikis bearbeitet werden, können auch als XML exportiert werden. Es ist aber noch nicht bekannt, ob diese Software die XML Daten auch in ein bestimmtes Schema, z.B. in ein RDF Schema, schreiben können. Falls dies nicht möglich sein sollte, so muss noch ein Transfer-Skript zur Übersetzung der Daten aus dem Semantic Media Wiki in das RDF Format zur Ingestierung in die Suchmaschine bereitgestellt werden.

Arbeitspunkt (6) Erstellung einer Abgleichmethode für die Wissensdaten

Sowohl die Daten des Tanzarchivs in der Akademie der Künste Berlin als auch die Daten des Tanzarchivs Leipzig werden (bzw. sind bereits) Teil der Datenverwaltung einer größeren Institution. Diese Institutionen müssten bei Aufgabe der lokalen Datenhaltung für Tanz-Wissen und die Tanz-Vokabularien ihre Mitarbeiter auf die Nutzung einer alternativen Datenbanktechnik (des ‚Share‘) einstellen, auch wenn diese Personen keine tanzthemenbezogene Daten pflegen müssen. Da diese Zentralisierung der Datenverwaltung nicht durchführbar ist, muss für die genannten Institute eine Alternative gefunden werden. Als Option bietet sich hier ein beidseitiger Datenabgleichprozess zwischen Tanz-Archiv und dem ‚Share‘ an.