

Denise Kiefer, Dr. Thomas Wollschläger

SHAMAN

Fortschrittsbericht zur Projekthalbzeit

Die Zeitschrift »Dialog mit Bibliotheken« berichtete bereits über das im März 2008 angelaufene EU-Forschungsprojekt SHAMAN (Sustaining Heritage Access through Multivalent Archiving) und gab einen ersten Überblick über die Inhalte und Ziele dieses Projektes.¹⁾ Zwei Jahre nach Start des Projektes wird an dieser Stelle ein Zwischenbericht gegeben, der die bisher absolvierten Schritte und neuen Erkenntnisse innerhalb des Projektes im Kontext der Langzeitarchivierung elektronischer Daten beschreibt.

Die Deutsche Nationalbibliothek (DNB), vertreten durch die Abteilung Informationstechnik, hat innerhalb des aus 18 Partnern bestehenden Konsortiums die Verantwortung für das Arbeitspaket »Dokumenterzeugung, Archivierung, Zugriff und Nachnutzung im Kontext von Gedächtnisorganisationen für wissenschaftliche und behördliche Sammlungen« übernommen. Sie ist dabei maßgeblich an der Entwicklung eines Prototyps beteiligt, welcher innovative Technologien, Strategien und Testmaterialien für ein modernes Langzeitarchivierungs-Rahmenkonzept zusammenführen wird. Eine grundlegende Voraussetzung für das Entwickeln einer international vernetzten Infrastruktur ist die Analyse vorhandener Systeme und institutioneller Herangehensweisen, Technologien und Archivierungsprozesse. Ziel ist, einen offenen und erweiterbaren Rahmen zu erzeugen, welcher alle Bestandteile, Services, Schnittstellen und Spezifikationen im Kontext der verbreiteten Langzeitarchivierungsstandards so definiert, dass ihre Wiederverwendung erleichtert wird. Die verteilte Archivierungsinfrastruktur wird unter Verwendung von GRID-Technologien²⁾ erstellt. SHAMAN entwickelt Konzepte, Technologien und Dienstleistungen, die prototypisch in Testumgebungen und in praktischen Szenarien ausgewertet werden.

Das gemeinsame Vorgehen zur Entwicklung eines solchen SHAMAN-Prototyps wurde im November

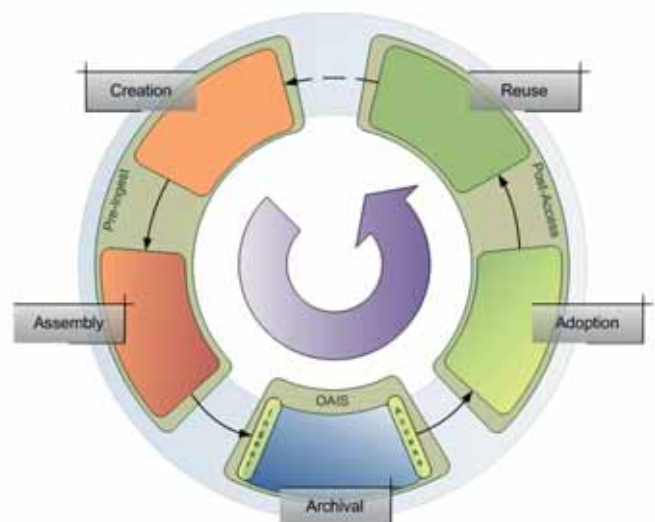
2009 innerhalb eines SHAMAN-Konsortialtreffens in Frankfurt am Main zwischen den Partnern abgestimmt. Während sich der Prototyp, für den die DNB verantwortlich ist, auf den Bereich der Gedächtnisorganisationen konzentriert, werden von anderen Projektpartnern zwei weitere Prototypen entwickelt, die sich auf digitale Objekte aus industriellem Design bzw. Produktionstechnik und auf Datenquellen aus e-Science Anwendungen beziehen.

In einem so genannten »Demonstrator«, welcher dem Aufbau des Prototyps für den Bereich Gedächtnisorganisationen vorausgeht, werden die verteilten Komponenten des angepeilten SHAMAN-Archivierungskonzepts aufgezeigt. Der Demonstrator verdeutlicht, wie eine bessere Infrastruktur gegenüber früheren Systemen aussehen kann und wie eine verbesserte und sicherere Archivierung des Archivmaterials der Gedächtnisorganisationen ermöglicht werden kann. Als Grundlage für den Demonstrator wiederum wurde von den Projektpartnern ein »Information Lifecycle ISP1« entwickelt (s. Abb.). Er beschreibt den Lebenszyklus elektronischer Publikationen aus der SHAMAN-Perspektive unter Verwendung und Berücksichtigung verschiedener Technologien und Softwarekomponenten.

SHAMAN-Konsortialtreffen im November 2009

Entwicklung eines »Demonstrators«

DNB entwickelt Prototyp für ein Langzeitarchivierungs-Rahmenkonzept



Der SHAMAN Information Lifecycle

Das Modell veranschaulicht exemplarisch die fünf verteilten Instanzen des SHAMAN-Archivierungskonzeptes. Diese Instanzen oder verschiedenen Phasen des Lifecycles sind:

- Creation (Produktion des Ausgangsmaterials),
- Assembly (Herstellung der Archivobjekte),
- Archival (Archivierung),
- Adoption (Übernahme in andere Archivsysteme) und
- Reuse (Nachnutzung).

Der Kreislauf beginnt mit der Produktion der Informationen, die als zu konservierende Objekte in den Kreislauf gelangen. Im besten Fall handelt es sich dabei nicht nur um die Objekte, sondern auch um die vom Produzenten mitgelieferten Kontextinformationen (z. B. Angaben zur Produktionsumgebung oder zu Datenbanken, die mit dem Objekt verknüpft sind). Die Erzeugung kontextsensitiver Metadaten ermöglicht eine bessere Wiederauffindbarkeit der Materialien und vereinfacht deren Nachnutzung.

In der Pre-ingest-Phase entsteht ein angereichertes Archivobjekt; die Anreicherungsphase wird durch den Prozess »Assembly« abgebildet. Hierbei wird die Generierung zusätzlicher Informationen für die Darstellung der Objekte (schließt sowohl beschreibende Metadaten als auch Langzeitarchivierungsmetadaten ein) und für deren Validierung adressiert. Danach werden so genannte »Submission Information Packages« (SIP) für die Einspeisung in das Archiv erzeugt. In dieser Phase wird eine Schnittstelle zu Bibliothekssystemen angeboten, um die bibliografische Indexierung von existierenden Katalogdaten aus Testsammlungen, die für das Demonstrationsszenario bereitgestellt wurden, zu simulieren.

Die Generierung von Langzeitarchivierungsmetadaten wird in der ersten Version des Prototyps unter Verwendung der JHOVE Software³⁾ demonstriert.

Die DNB verfügt mit der im Rahmen des Projektes kopal⁴⁾ entwickelten koLibRI-Software über eine Lösung, die in einer integrierten Methode unter Nutzung der genannten Software die gewünschten Metadaten generiert.

Die Archivierungsphase beginnt mit Aufnahme der SIPs in das GRID-basierte Archivsystem und berücksichtigt darüber hinaus die Datenverwaltung. Bestandserhaltende Maßnahmen werden für den ersten Schritt des Prototyps vorerst ausgeklammert. Bevor die Archivobjekte nachgenutzt werden können, müssen sie bei Abruf aus dem Langzeitarchiv in einer zusätzlichen Phase so angepasst werden, dass sie von unterschiedlichen Zielsystemen nachgenutzt werden können. Dazu werden fortgeschrittene Such- und Ausgabefunktionalitäten zur Verfügung gestellt. Diese Funktionalitäten stehen nicht im Fokus des SHAMAN-Prototyps, können aber leicht integriert werden.

Damit sind essenzielle Bauteile eines ersten digitalen Archiv- und Bestandserhaltungsprototyps für das Demonstrationsszenario für die Gedächtnisorganisationen vorhanden. In der zweiten Projekthälfte werden die Partner auf Basis dieses Demonstrators einen stabilen Prototyp entwickeln.

Zur Vorbereitung des im April 2010 stattfindenden Reviews durch die beauftragten Experten der zuständigen EU-Kommission fand im Februar ein Pre-Review in Frankfurt am Main statt. In diesem wurde die Vorstellung des entwickelten Demonstrators vor den Partnern aus dem SHAMAN-Konsortium erprobt und evaluiert.

In der DNB werden derzeit verschiedene Technologien, welche von den Partnern in den Demonstrator eingebracht worden sind, eingehend untersucht. Ziel dabei ist es, Synergie- und Nutzungspotenzial für Prozesse und Strategien der DNB zu ermitteln.

Produktion der
Informationen

Anreicherungs-
phase der
Objekte

Archivierungs-
phase

Perspektive

Anmerkungen

1 Vgl. hierzu auch: Altenhöner, Reinhard; Kranstedt, Alfred: SHAMAN. In: Dialog mit Bibliotheken, 20 (2008) 2, S. 29 – 34.

2 Unter GRID-Computing versteht man eine verteilte Rechnerinfrastruktur, wobei ein virtueller Hochleistungsrechner aus einem Cluster lose gekoppelter Einzelcomputer besteht.

3 JHOVE=JSTOR/Harvard Object Validation Environment;
JHOVE-Website: <<http://hul.harvard.edu/jhove/>>

4 Wollschläger, Thomas: kopal goes live. In: Dialog mit Bibliotheken, 19 (2007) 2, S. 17 – 22. kopal-Website:
<<http://kopal.langzeitarchivierung.de/>>