

Slot 6.3: Digitale Langzeitarchivierung

Moderation: Bettina Kann (*Österreichische Nationalbibliothek, Wien, Österreich*)

Data Curation oder was kostet uns die digitale Serviceleistung wirklich. Ansätze und Erfahrungen im Münchner Digitalisierungszentrum/Bayerische Staatsbibliothek (20min)

Vortragende/r: Kempf, Klaus (*Bayerische Staatsbibliothek, München, Deutschland*)

Zur Person: Leiter der Hauptabteilung Bestandsaufbau und Erschließung, insbesondere Digitale Bibliothek.

National und international seit Jahren mit strategischen Fragestellungen im Bibliotheksmanagement befasst. Mit dem Digitalisieren analoger Materialien und der (Einmal)Präsentation der digitalen Ergebnisse im Internet ist es nicht getan. Diese Erfahrung machen zunehmend alle Bibliotheken, die sich dem komplexen Thema Digitalisierung und dem Aufbau entsprechender Serviceleistungen intensiver widmen. Insbesondere wenn kritische Datenmengen erreicht werden und die Hard- und Softwareausstattung dann in neue Größenordnungen hineinwachsen muss sowie die Wartung und Pflege der Daten bzw. der dazu gehörigen Metadaten und der für ihre Bereitstellung und Archivierung benötigten Softwaretools (Stichwort: Migration) zu erheblichen Personal- und Finanzaufwänden führen, stellt sich die Frage nach einem stimmigen Data curation-Konzept. Spätestens jetzt zeigt es sich, dass die digitale Bibliothek eine erhebliche personell-finanzielle Herausforderung auf Dauer für die Bibliothek darstellt und entsprechende aufbau- und ablauforganisatorische Vorkehrungen notwendig sind. Die BSB hat auch bei diesem Thema mittlerweile reiche einschlägige Erfahrungen sammeln und entsprechende Lösungsansätze entwickeln können, die auch für andere Bibliotheken bzw. Gedächtniseinrichtungen ganz allgemein von großem Interesse sein können. In dem Vortrag werden sie in wesentlichen Zügen vor- und zur Diskussion gestellt.

Open Community Approaches to Digital Preservation (20min)

Vortragende/r: King, Ross (*Austrian Institute of Technology, Wien, Österreich*)

Zur Person: Senior Scientist beim Austrian Institute of Technology. Forschungsinteressen: Verwaltung, Suche und Langzeiterhaltung von multimedialen Inhalten. Chairman of the Board der Open Preservation Foundation (non-profit, Bereich digitale Langzeitarchivierung).

Libraries and archives face an increasing shift towards the long-term management of digital cultural heritage and requirements that new types of content remain accessible for future generations. Community collaboration to develop open and shared solutions is a necessary response to the challenges, but requires active engagement from institutions of all sizes to consolidate learning and build trust in proven approaches. Current state-of-the-art will be presented based on recent community analysis along with future trends and opportunities.

Langzeitarchivierung von Webseiten: Herangehensweise und Praxistext mit PABLO (20min)

Vortragende/r: Herschung, Alexander (*Startext GmbH, Bonn, Deutschland*);

Bargmann, Monika (*Wienbibliothek im Rathaus, Wien, Österreich*)

Zu den Personen: Alexander Herschung ist Geschäftsführer der Startext GmbH; Monika Bargmann leitet das Grüne Archiv und arbeitet im Bereich Digitale Services in der Wienbibliothek im Rathaus.

Im Kontext der Webarchivierung kann vor allem das Problem der Langzeitarchivierbarkeit der beteiligten Dateiformate als offenes Problem gelten. Im Internet publiziertes Material ist häufig flüchtig. Die Ansprüche an Webauftritte ändern sich schnell und zugleich in voneinander unabhängigen Dimensionen: fachlich/inhaltliche

Anforderungen, technische Anforderungen, Darstellungs-/Designanforderungen. Im Ergebnis sind Webseiten häufigen Relaunches unterworfen. Die Vorversionen, obgleich für die historische Betrachtung von Interesse, gehen dabei häufig verloren.

Die einfache Speicherung der Webseiten als offline-Version ist keine dauerhafte Lösung. Durch die Weiterentwicklung der darstellenden Browser und zugrunde liegenden Betriebssysteme verändert sich die Darstellung der Webseite auch dann, wenn der zu Grunde liegende Code (html + zugehörige Dateien) unverändert bleibt. Die Herausforderung ist also: Wie kann das Wesentliche der Webseite, d.h. die Art und Weise, wie sie „heute“ durch einen „bestimmten Browser“ „dargestellt“ wird und „bedienbar“ ist, dauerhaft bewahrt werden? Eine etablierte Lösung zu diesem Problem existiert bisher nicht.

Die Firma startext hat mit PABLO kürzlich ein Werkzeug erstellt, mit dem sich das Format von Webseiten zum Zwecke der dauerhaften Archivierung drastisch vereinfachen lässt. PABLO versucht, einen Lösungsansatz bereitzustellen, indem das zu speichernde Datenformat radikal vereinfacht wird. PABLO erzeugt aus der Vielzahl von Ursprungsdateien lediglich zwei Dateitypen, aus denen sich aber die „Haptik“ der ursprünglichen Webseite wiederherstellen lässt: die so genannte Archivform, die aus einer Bilddatei, die die Darstellung im Browser speichert, und einer METS-XML-Datei, die im Wesentlichen Position und Ziel der auf der Seite enthaltenen Verlinkungen enthält, besteht. Aus dieser Archivform kann mit PABLO eine Präsentationsform erzeugt werden, die in Darstellung und Verhalten der ursprünglichen Webseite nahekommt. Die Präsentationsform verfügt über navigierbare Links und kann von den AnwenderInnen somit „gesurft“ werden. Diesen Ansatz kann man sicher kontrovers diskutieren. Wir würden ihn gerne offen vorstellen, um entsprechende Diskussionen anzuregen.

Dargestellt wird die Art und Weise der Erzeugung der Archivform (rekursives Durchlaufen der Webseite in der vorgegebenen Tiefe) sowie die erzeugte Datenstruktur (hier insbesondere die Struktur und Inhalte der METS-XML-Dateien). Momentan bestehende Begrenzungen und mögliche zukünftige Weiterentwicklungen werden thematisiert.

Ergänzt wird dieser Einblick durch einen Praxistest. Das Grüne Archiv, eine kleine Informationseinrichtung, steht vor der Herausforderung, sehr umfangreiche und komplexe Websites mit verschiedenen Dokumententypen archivieren zu müssen. Es wird offengelegt, wie die Archivierung mit Pablo – im Gegensatz zu einer bloßen Archivierung der Website-Dateien – funktioniert, wo sie bei den getesteten Websites gut funktioniert hat und in welchen Bereichen Einschränkungen aufgetreten sind.

DIPS – Erfahrungen aus der Implementierung eines neuen digitalen Langzeitarchivsystems (20min)

Vortragende/r: *Kann, Bettina (Österreichische Nationalbibliothek, Wien, Österreich)*

Zur Person: Studium der Geschichte in Wien, Ausbildung zur Bibliothekarin 1998. Seit 1993 an der Österreichischen Nationalbibliothek tätig, seit 2008 Leiterin der Hauptabteilung Digitale Bibliothek, unterrichtet im ULG und ALG „Library and Information Studies“.

Die Österreichische Nationalbibliothek hat 2014 nach erfolgter Ausschreibung ein neues Archivierungssystem für digitale Medien angeschafft. Dieses war notwendig geworden, da das alte Digital Asset Management System an die Grenze seiner Leistungs- und Skalierfähigkeit angelangt war. Der Vortrag beschreibt Konzept und Architektur der gesamten Infrastruktur zur digitalen Langzeitarchivierung an der Österreichischen Nationalbibliothek und erläutert die wesentlichsten Funktionalitäten im neuen Archivsystem. Weiters fokussiert der Vortrag auf dem Migrationsprozess der unterschiedlichen Daten aus dem alten in das neue System und beleuchtet die „lessons learnt“ aus dem Projekt.