

Praktikum an der Library of Congress

Office of Strategic Initiatives, NDIIPP

6. April – 15. Mai 2009

Sabine Schrimpf, Deutsche Nationalbibliothek

Hintergrund.....	2
Library of Congress, Office of Strategic Initiatives, NDIIPP	2
Planungen für die Nachhaltigkeit von NDIIPP.....	2
„Communicating Digital Preservation“ – Arbeit mit dem Communications Team.....	4
Web Archiving an der Library of Congress.....	4
Besuch des National Audio-Visual Conservation Center	5
Teilnahme an der <i>Archiving 2009</i>	6



Hintergrund

Vom 6. April bis zum 15. Mai 2009 absolvierte ich ein Praktikum am Office of Strategic Initiatives der Library of Congress. Dort arbeitete ich mit dem Team, welches das *National Digital Information Infrastructure and Preservation Program* (NDIIPP) koordiniert, sowie mit dem Web Archiving Team der Library of Congress. Der Aufenthalt wurde von BI-International finanziell unterstützt.

Das Praktikum war Bestandteil des weiterbildenden Masterstudiengangs Bibliotheks- und Informationswissenschaft per Fernstudium an der Humboldt-Universität zu Berlin, den ich neben meiner Tätigkeit in *nestor* (Kompetenznetzwerk digitale Langzeitarchivierung) an der Deutschen Nationalbibliothek in Frankfurt absolviere. Aus dem *nestor*-Kontext war mir die allgemeine Ausrichtung von NDIIPP schon bekannt und ich war daran interessiert, einen tieferen Einblick in die Arbeit des Programms zu bekommen. In einer Phase, in der *nestor* und NDIIPP den Übergang von Projektorganisation in nachhaltige Organisationsformen vorbereiten, interessierten mich besonders die Pläne der Library of Congress, die NDIIPP-(Projekt-)Partnerschaften zu formalisieren. Sie werden in diesem Bericht als erstes vorgestellt. Anschließend gehe ich auf die weiteren Aktivitäten ein.

Während meiner Praktikumszeit wurde ich außer in allgemeine NDIIPP-Aktivitäten in den Arbeitsalltag von zwei kleineren Arbeitsgruppen eingebunden: das *NDIIPP Communications Team* (Öffentlichkeitsarbeit, Newsletter, Webseite) und das *Library of Congress Web Archiving Team*. Daneben besuchte ich das *National Audio-Visual Conservation Center* der Library of Congress in Culpeper, Virginia, und die *Archiving 2009 Konferenz* in Arlington, Virginia.

Library of Congress, Office of Strategic Initiatives, NDIIPP

Die Library of Congress ist die Referenzbibliothek des amerikanischen Kongresses und de facto die Nationalbibliothek der Vereinigten Staaten. Mit über 32 Millionen Büchern und anderen Printmaterialien und mehr als 61 Millionen Manuskripten ist sie eine der größten Bibliotheken der Welt. Durch das United States Copyright Office, das an der Library of Congress angesiedelt ist, erhält die Bibliothek eine Kopie eines jedes Druckes oder Musikträgers, für den in den Vereinigten Staaten ein Copyright beantragt wird.

Das *Office of Strategic Initiatives* unterstützt die Strategieplanung der Library of Congress, besonders im Hinblick auf die Positionierung der Bibliothek in einer digitalen Welt. Es koordiniert unter anderem die Webservices der Bibliothek und das NDIIPP-Programm.

Das Ziel von NDIIPP ist es, eine nationale Strategie zur Sammlung und Langzeitarchivierung von digitalen Ressourcen zu entwickeln. Die Bibliothek wurde 2000 vom amerikanischen Kongress mit der Durchführung des Programms beauftragt. Seitdem hat die Library of Congress mit Partnern aus den ganzen USA am Aufbau eines Partner-Netzwerks sowie einer technischen Langzeitarchivierungs-Infrastruktur gearbeitet.

Planungen für die Nachhaltigkeit von NDIIPP

Mit Projektmitteln gestartet, verfügt NDIIPP 2009 zum ersten Mal über ein reguläres Budget von \$ 6 Mio, die als Extraposten im Haushalt der Library geführt werden. Ursprünglich hatte der amerikanische Kongress der Library im Jahr 2000 ein Budget von insgesamt \$ 100 Mio für das NDIIPP-Programm zur Verfügung gestellt. Davon wurden ca.

50% bis 2007 für die erste Runde von Projekten ausgegeben. Doch dann forderte der in dem Jahr neu zusammengetretene Kongress sämtliche auf dem NDIIPP-Konto verbliebenen Mittel zurück. Mit Restmitteln aus laufenden Projekten konnte NDIIPP in 2007 dennoch zwei große Projekte anstoßen: Die *State Government Initiative* (<http://www.digitalpreservation.gov/partners/states.html>) und *Creative America* (<http://www.loc.gov/today/pr/2007/07-156.html>). Im Jahr 2008 musste das NDIIPP-Team mit dem Congress ein neues Budget verhandeln, mit dem Resultat, dass NDIIPP ab 2009 als ein Posten im Haushalt der Library of Congress mitgeführt und mit einem jährlichen Budget von \$ 6 Mio ausgestattet wird – auf vorerst unbegrenzte Zeit. Das gibt den Partnern Planungssicherheit für die Zukunft.

Die Library of Congress erarbeitet derzeit eine *National Digital Preservation Charter*, unter der sich ein großer Teil der NDIIPP-Partner zu einem Netzwerk zusammenschließen und das eine dezentrale nationale Sammlung digitaler Ressourcen anlegen soll. Das NDIIPP-Management an der Library of Congress möchte die Charter noch 2009 in Kraft setzen.

Es ist vorgesehen, dass aus dem NDIIPP-Budget weiterhin Langzeitarchivierungsprojekte der NDIIPP-Partner finanziert werden sollen. Partnerprojekte wurden bislang immer nach dem gleichen Muster angestoßen. Dieses Prozedere soll auch in Zukunft beibehalten werden:

Wenn ein Thema identifiziert ist (z.B. *State Government Information* oder *Basic Technical Infrastructure, Tools and Services for the NDIIPP network*), lädt die Library of Congress die relevanten Stakeholder zu einer *Convening Session* (eine Art Auftaktworkshop) ein. Innerhalb von sechs Monaten folgt ein *Request for Expressions of Interest* (siehe z.B. für die *State Government Initiative* http://www.digitalpreservation.gov/partners/pdf/states_rfei_final.doc). Im Fall der *State Government Initiative* konnte die Library vier Pilotprojekte unter 14 eingereichten Projektvorschlägen auswählen.

Ein Schwerpunkt für die mittelfristige Zukunft ist der Aufbau einer nationalen digitalen Sammlung. Für 2009 sind drei *Convening Sessions* geplant, jeweils zu einer nationalen Sammelstrategie von a) *digital news and journalism*, b) *geospatial content*, c) *web harvesting*. Die ersten beiden Themen lassen sich in den Vorschlag eines *National Digital Collection Frameworks* einordnen, den die Library of Congress gemeinsam mit NDIIPP-Partnern Anfang 2009 erarbeitet hat. Der *National Digital Collection Framework* sieht zehn Themenbereiche vor:

- American History
- Arts and Culture
- Government, Politics and Law
- Maps and Geography
- News, Media and Journalism
- Religion and Philosophy
- Social Sciences
- Science, Mathematics and Technology
- World History and Cultures
- Sports, Recreation and Travel

In allen NDIIPP-Projekten sind die Partner für die Speicherung und Langzeitarchivierung der Sammlungen, die im Rahmen der Projekte entstehen, selbst verantwortlich. Der Entwurf der *National Digital Preservation Charter* sieht vor, dass die Partner allen anderen Partnern, mindestens aber der Library of Congress, Zugriffsrechte auf die digitalen Sammlungen einräumen, so dass eine wirkliche verteilte nationale Sammlung digitaler Ressourcen entsteht.

„Communicating Digital Preservation“ – Arbeit mit dem Communications Team

Öffentlichkeitsarbeit ist ein wichtiger Bestandteil des NDIIPP-Programms. Sie richtet sich in diesem Fall im Wesentlichen an vier Zielgruppen: die NDIIPP-Partner, weitere Fachleute aus der Digital Preservation Community, Entscheidungsträger und die allgemeine Öffentlichkeit. Als Distributionskanäle werden genutzt:

- Die Programmhauptseite <http://digitalpreservation.gov/>
- Ein monatlicher Newsletter und RSS feeds
- „Briefing Books“ für Kongressabgeordnete
- Artikel in Fachzeitschriften
- weitere Publikationen
- Konferenzbeiträge
- Broschüren und Flyer
- Videos

Im Mittelpunkt der Öffentlichkeitsarbeitsstrategie steht die Homepage <http://digitalpreservation.gov>. Sie bündelt Informationen zu allen NDIIPP-Aktivitäten sowie zu Projekten und beteiligten Partnern. Publikationen, Präsentationen und Aufzeichnungen von Konferenzbeiträgen werden zum Download bereitgehalten. In einer Liste von Tools und Services werden Ergebnisse aus NDIIPP-Projekten zusammengetragen.

Zwei Bereiche der Webseite sollen in 2009 ausgebaut werden: Ein „Personal Archiving“ Bereich richtet sich an Privatanwender. Angeboten werden dort einführende Informationen zum Thema und Tipps zur sicheren Aufbewahrung persönlicher digitaler Dokumente (direkter Zugang: <http://www.digitalpreservation.gov/you/>). Die Seite gehört seit einigen Monaten zu den am häufigsten aufgerufenen innerhalb der digitalpreservation.gov Präsenz. Das Informationsangebot ist allerdings noch unvollständig. Dem großen Interesse entsprechend soll die Seite benutzerfreundlicher werden und ein größeres Informationsangebot, strukturiert nach Dokumentarten (Digital Photographs, Digital Video, Digital Audio, Text Documents, Websites und E-Mail) bereitgestellt werden. Die Informationen sollen ohne technische Vorkenntnisse verstanden werden können. Eine vorherige Recherche hatte ergeben, dass es bislang kein solches Informationsangebot im Internet gibt. Vorhandene Informationsangebote richten sich größtenteils an Institutionen und setzen entweder umfassendes IT-Wissen oder IT-Kapazitäten voraus.

Der zweite Bereich der Internetpräsenz digitalpreservation.gov, der ausgebaut werden soll, sind Videoclips zum Thema Digital Preservation. Auf der Homepage gibt es bereits einige Mitschnitte von Vorträgen, die von NDIIPP-Mitgliedern gehalten wurden. Das Angebot soll um speziell für die Seite produzierte Info-Clips erweitert werden, die in kurzer und prägnanter Form die Highlights aus dem Programm vorstellen.

Web Archiving an der Library of Congress

Anders als die Deutsche Nationalbibliothek hat die Library of Congress keinen gesetzlichen Sammelauftrag für digitale Publikationen. Ihrem traditionellen Auftrag entsprechend, Materialien von historischer Bedeutung für den amerikanischen Congress und das amerikanische Volk zu sammeln, baut die Library of Congress dennoch digitale Sammlungen auf. Alle öffentlich zugänglichen Websammlungen der Library können unter <http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html> abgerufen werden.

Seit 2000 hat die Library thematische Webarchive unter anderem zu den United States National Elections of 2000, 2002, and 2004, dem Irak-Krieg, und den Ereignissen des 11. September 2001 angelegt. Seit 2004 gibt es für diese Aufgabe ein eigenes *Web Capture*

Team. In Absprache mit den Fachabteilungen wählt das Team Sammlungen und Inhalte für die jeweiligen Sammlungen aus. Das Harvesting wird für die Library vom *Internet Archive* durchgeführt. Das *Web Capture Team* überwacht den Sammelprozess, koordiniert die darauf folgende Erschließung und ist für die Qualitätskontrolle zuständig. Im Folgenden wird ein grober Überblick über den Workflow für Websammlungen an der Library of Congress gegeben:

1. Vorschlag – Eine Fachabteilung der Library schlägt eine thematische Sammlung vor.
2. Zustimmung – Ein eigenes *Web Archiving Management Oversight Committee*, das aus Senior Managern verschiedener Abteilungen zusammengesetzt ist, prüft den Vorschlag und stimmt zu, wenn er mit den Sammelrichtlinien der Library übereinstimmt.
3. Information – Wenn die Zustimmung erteilt ist, erbittet das *Web Capture Team* detaillierte Informationen über die geplante Websammlung von der zuständigen Fachabteilung.
4. Auswahl der Webseiten und Rechtlklärung – Gemeinsam mit der Fachabteilung wählt das *Web Capture Team* Webseiten für die Sammlung aus und beginnt mit der Rechteeinholung. Für jede Webseite muss der Rechteinhaber identifiziert und kontaktiert werden. Stimmt der Rechteinhaber der Sammlung nicht ausdrücklich zu, darf die Seite nicht gesammelt werden.
5. Dokumentation – Für jede Websammlung wird eine interne Projektseite kreiert, auf der alle Informationen und der Status der ausgewählten Webseiten dokumentiert werden.
6. Seed List – Die URLs aller ausgewählten Webseiten werden zu einer Seed List zusammengestellt. Diese Seed List ist Grundlage des Crawls (siehe Punkt 7. unten), den das Internet Archive für die Library ausführt.
7. Crawl – Das *Internet Archive* fertigt Kopien der ausgewählten Seiten an und das *Web Capture Team* führt laufend Qualitätskontrollen durch.
8. Nachbearbeitung und Erschließung – Die Katalogisierungsabteilung erschließt die gesammelten Webseiten nach dem *Metadata Object Description Schema* (MODS).

Besuch des National Audio-Visual Conservation Center

Der Packard Campus des *National Audiovisual Conservation Center* (NAVCC) in Culpeper, Virginia, ist das jüngste Gebäude der Library of Congress. Er ist 60 Meilen (1,5 Autostunden) von den Hauptgebäuden der Library of Congress in Washington, D.C., entfernt, doch die Vorzüge des Gebäudes rechtfertigen die verkehrstechnisch ungünstige Lage. Bis 1988 ein Hochsicherheitsgebäude der Federal Reserve Bank, wurde der Bau 1997 von David Packard gekauft und der Library of Congress für ihre audiovisuellen Sammlungen geschenkt. Mit großzügiger finanzieller Unterstützung des Packard Humanities Institutes wurde der Bau in den folgenden zehn Jahren für die speziellen Bedürfnisse der Audio- und Filmsammlungen der Library of Congress umgerüstet und 2008 in Betrieb genommen.

Auf dem Campus werden Arbeitsstationen und Maschinen für die Konservierung und Konvertierung von analogen audiovisuellen Medien (Videokassetten, Fotofilm, Videofilm, Audiokassetten, Schallplatten etc.) vorgehalten. Während die analogen Techniken noch den größten Raum auf dem Campus einnehmen, sind auch enorme Kapazitäten für Digitalisierung und digitale Speicherung vorgesehen. Das digitale Archivsystem ist darauf ausgerichtet, pro Jahr mehrere Petabyte (10^6 Gigabyte) audiovisueller Ressourcen aufzunehmen. Das Langzeitarchivierungskonzept sieht kontinuierliche Migration sowie ständige Datenüberprüfung vor.

Langfristig ist geplant, sämtliche Ton- und Videoaufnahmen der Library of Congress zu digitalisieren. Daher wird derzeit ein Pilotprojekt im Bereich der automatischen Digitalisierung mit 3/4" Videokassetten durchgeführt (SAMMA¹). Die digitalisierten Kopien sollen über eine High-Speed-Internetverbindungen unmittelbar nach der Digitalisierung den Nutzern in den Lesesälen der Library of Congress in Washington, D.C., zugänglich gemacht werden. Auch Digitalisierung on Demand anstelle von Dokumentenlieferung in die Lesesäle ist vorgesehen.

Teilnahme an der *Archiving 2009*

Die Konferenz fand vom 4.-7. Mai in Arlington, Virginia, statt (<http://www.imaging.org/conferences/archiving2009/>). Das Konferenzprogramm beinhaltete die Themenblöcke: Digital Collection Stewardship und Imaging and Preservation. Im ersten Programmblock wurden Themen behandelt wie die wirtschaftliche Nachhaltigkeit von Langzeitarchivierung, Selektions- und Archivierungsstrategien für Wissenschaftsdaten, Ingest-Anforderungen, Risikoanalysemethoden für digitale Sammlungen und die Verbreitung von PREMIS in Gedächtnisorganisationen. Gleich drei Vorträge stellten Projekte zur Langzeitarchivierung von Erdbeobachtungsdaten (geospatial data) vor, von denen zwei im Rahmen von NDIIPP gefördert werden. Insgesamt beschäftigten sich vier Vorträge und sechs Posterpräsentationen mit NDIIPP-Aktivitäten.

Im Themenblock Imaging and Preservation dominierten Vorträge zu Digitalisierungstechnologien und –workflows. Am zweiten und dritten Tag klangen die Konferenztage jeweils mit interaktiven Poster Sessions zu Themen wie Suchstrategien für Webarchive, „Digitale Archäologie“ und Archivierungsstrategien für digitale Fotos aus.

¹ System for the Automated Migration of Media Archives. Aus der Projektbeschreibung: “Over the next several years the Library will use the SAMMA robotic cassette reformatting system to migrate and digitize its 3/4" video holdings, and will eventually expand to include other cassette-based video and audio formats. The final product will be preservation-quality MJPEG2000 digital files and the technical metadata describing the condition of the media item and the migration process.