



**World Library and Information
Congress:
70th IFLA General Conference and
Council**

**22-27 August 2004
Buenos Aires, Argentina**

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

Code Number:	026-E
Meeting:	148. IFLA-CDNL Alliance for Bibliographic Standards (ICABS)
Simultaneous Interpretation:	-

Sammlung und Verwaltung von Web-Ressourcen für den Langzeit-Zugriff: Web-Harvesting und Richtlinien für die Erhaltung (ICABS actions 3.3 und 3.4)

Pam Gatenby

Assistant Director General
Collection Management
National Library of Australia
pgatenby@nla.gov.au

Abstract:

Die National Library of Australia leistet Ihren Beitrag zum ICABS-Arbeitsplan durch das Ziel Nummer 3: Vertiefung des Verständnisses von Themen, die mit der Langzeitarchivierung elektronischer Ressourcen zusammenhängen. Sie hat sich zu Maßnahmen verpflichtet, die die Zusammenarbeit in den Bereichen Web-Archivierung (Action 3.3) und Erhaltung digitaler Ressourcen (Action 3.4) fördern, wobei sie ihre eigenen Erfahrungen auf diesen Gebieten einbringt. Zu den spezifischen Aktivitäten gehören die Leitung der Deep Web Working Group des International Internet Preservation Consortium (IIPC), die Veranstaltung einer internationalen Konferenz über Web-Archivierung, Arbeiten zur Automatisierung der Speicherung und Archivierung von Online-Amtsdruckschriften, die Verbreitung von Informationen über Vorgehensweisen bei der Web-Archivierung und das Sammeln vorhandener Informationen, um Hilfestellung bei Entscheidungen im Zusammenhang mit der Erhaltung digitaler Ressourcen zu geben. Ziel ist es hierbei, Informationen zu diesen Themen leichter verfügbar und über das Subject Gateway PADI (<http://www.nla.gov.au/padi/>) bequem zugänglich zu machen.

1. Einführung

Die National Library of Australia freut sich, Teilnehmer der ICABS-Allianz zu sein, da wir uns zur Zusammenarbeit bei der Entwicklung von Standards und Strategien verpflichtet haben, die darauf ausgerichtet sind, den kostengünstigen Zugang zu den Sammlungen von nationalem Erbe zu unterstützen.

Unser Hauptbeitrag zum ICABS-Arbeitsprogramm liegt in Ziel 3: Vertiefung des Verständnisses von Themen, die mit der Langzeitarchivierung elektronischer Ressourcen zusammenhängen.

Wir wollen zu diesem speziellen Ziel beitragen, weil wir Erfahrung mit digitaler Archivierung haben, und weil uns bewusst ist, dass viele Nationalbibliotheken Anleitung in diesem Bereich suchen, da sie unsicher sind, wie sie vorgehen sollen. Wir haben auch erkannt, dass es zu unserer eigenen, ständiger Weiterentwicklung notwendig ist, in praxisbezogener Partnerschaft mit anderen zusammenzuarbeiten, da es zu viele Fragen gibt, um sie alleine lösen zu können.

ICABS hat uns einen Anstoß gegeben, intensiv darüber nachzudenken, wie wir unser Wissen und unsere Erfahrung effektiver mit anderen austauschen können, und wie wir sinnvoller zu dem Ziel der Langzeitverfügbarkeit digitaler Ressourcen von nationaler Bedeutung beitragen können.

2. Hintergrund zu den Aktivitäten der National Library of Australia im Bereich der digitalen Archivierung und Erhaltung

Digitale Archivierung

Seit 1996 sammelt die National Library of Australia wichtige australische Web-Sites und verwaltet sie in PANDORA, dem Webarchiv Australiens (<http://www.nla.gov.au/pandora/>). Dies ist mittlerweile eine Routinetätigkeit im Rahmen des Bestandsaufbaus, die von einem Team aus fünf Mitarbeitern erledigt wird. Die Titel werden mit Zustimmung der Verlage gesammelt, weil das australische Pflichtexemplargesetz elektronische Publikationen noch nicht abdeckt.

Das Archiv beinhaltet zurzeit rund 6.000 Titel und mehr als 10.000 Schnappschüsse (oder Zeitschnitte) von Web-Sites und ist über ein halbes Terabyte groß. Die Entwicklung von PANDORA ist ein Gemeinschaftsunternehmen, an dem die australischen Staatsbibliotheken und drei andere Institutionen mit nationalen Sammlungen beteiligt sind. Die Auswahl der Web-Sites basiert auf detaillierten Kriterien mit Schwerpunkt auf den Wert für die Forschung und den Informationsgehalt ebenso wie auf die kulturelle Bedeutung, und die auch einen Rahmen für eine Selektion nach Fächern bieten. Jede archivierte Web-Site ist qualitätsgeprüft im Hinblick auf Inhalt und Funktionalität.

Zur Verwaltung und Unterstützung ihrer Verantwortlichkeiten im Bereich der digitalen Archivierung hat die Bibliothek ihr eigenes digitales Archivierungssystem entwickelt, das unter dem Namen PANDAS (Pandora Digital Archiving System) bekannt ist. Das System unterstützt folgende Funktionen:

- + Verwaltung von Metadaten zu ausgewählten und abgelehnten Titeln
- + Beginn der Sammlung ausgewählter Titel
- + Ausführung des Qualitätsprüfungsprozesses

- + Vorbereitung von Dokumenten für die öffentliche Bereitstellung im Archiv
- + Verwaltung der Zugriffsrechte
- + Persistent-Identifizier-Verwaltung
- + Bereitstellung von Management-Reports

Mehrere Institutionen aus der ganzen Welt haben angefragt, ob sie PANDAS benutzen dürfen, weil es auf dem Markt bislang kein vergleichbares kommerzielles System gibt, das die Sammlung und Verwaltung von Web-Ressourcen unterstützt. Daraufhin hat die Bibliothek PANDAS auf Anforderung für Evaluierungszwecke zur Verfügung gestellt. Unter der Voraussetzung, dass wir keinen Support leisten, sind wir auch darauf eingestellt, anderen die Software zur Anwendung zur Verfügung zu stellen.

Digitale Erhaltung

Von Anfang an waren Überlegungen und Strategien zur Erhaltung integraler Bestandteil beim Aufbau digitaler Sammlungen. Ziel bei der Sammlung von Web-Ressourcen, die in PANDORA archiviert werden, ist es, diese dauerhaft zu verwalten, um den permanenten öffentlichen Zugriff darauf anzubieten. Dies erfordert sowohl kurz- als auch langfristige Planung und aktive, tägliche Bearbeitung der archivierten Ressourcen. Die digitale Archivierungspolitik (<http://www.nla.gov.au/policy/digpres.html>) und der Aktionsplan beschreiben die Hauptrichtungen und Aktivitäten, die von der Bibliothek verfolgt werden. Schwerpunkte sind gelegt auf:

- + Datenmanagement (Refreshing, Back-Ups, etc.)
- + Identifizierung und Dokumentation der Charakteristika archivierter Ressourcen
- + Identifizierung und Bewältigung von Risiken und Gefahren, die mit verschiedenen Dateiformaten verbunden sind
- + Entwicklung von Erhaltungsstrategien für jedes Dateiformat
- + Benutzung von Metadaten zur digitalen Langzeiterhaltung
- + Planung zur Notfallvorsorge

Kooperation ist ebenfalls zentraler Bestandteil der Herangehensweise der Bibliothek an die digitale Erhaltung. Um die Zusammenarbeit zu erleichtern, pflegt die Bibliothek das Subject Gateway PADI (Preserving Access to Digital Information, <http://www.nla.gov.au/padi>). PADI enthält Links zu einem breiten Spektrum an Informationen zur digitalen Erhaltung und bietet Hintergrundinformationen zu Schlüsselthemen. PADI ist ein zentraler Bestandteil unseres Beitrags zu ICABS und ermöglicht uns, den Informationsaustausch zu maximieren.

3. Die Beteiligung der National Library of Australia an ICABS

Die National Library of Australia leistet ihren Beitrag zum ICABS Action Plan in zwei Bereichen – durch Web-Harvesting (Action 3.3) und durch Langzeiterhaltung digitaler Materialien (Action 3.4)

Web-Harvesting (ICABS Action 3.3)

Ziel unserer Aktivitäten in diesem Bereich ist es, Methoden für die Sammlung und Archivierung von Netzpublikationen zu erforschen und zu fördern sowie damit verwandte Problemfelder zu erkennen. Dies geschieht zu dem Zweck, diese Informationen weit zu verbreiten und über PADI leicht zugänglich zu machen mit dem speziellen Ziel, denjenigen Institutionen zu helfen, die digitale Archivierungs- und Langzeiterhaltungsprogramme planen oder überprüfen.

Schlüsselaktivitäten sind:

- + Leitung der Arbeit der Working Group on the Deep Web des International Internet Preservation Consortium (IIPC)
- + Durchführung einer internationalen Konferenz über die Archivierung von Web-Ressourcen
- + Automatisierung der Speicherung und Archivierung von Online-Amtsdruckschriften und
- + Erforschung und Förderung von Methoden der Archivierung von Netzpublikationen, die durch Web-Harvesting gesammelt wurden.

(i) Leitung der Arbeit der Working Group on the Deep Web des International Internet Preservation Consortium (IIPC)

Das International Internet Preservation Consortium (IIPC) wurde 2003 gegründet, um die Zusammenarbeit zwischen Nationalbibliotheken bei der Entwicklung von Standards und Tools zur Unterstützung der Archivierung und Langzeiterhaltung von Web-Ressourcen zu erleichtern. Elf Nationalbibliotheken und das Internet Archive sind Mitglieder des Konsortiums, das von der Bibliothèque nationale de France koordiniert wird. Der Schwerpunkt des Konsortiums ist praxisbezogen. Die Teilnehmerbibliotheken zahlen einen Mitgliedsbeitrag und sind verpflichtet, aktiv zu dem Arbeitsprogramm der Gruppe beizutragen.

Die National Library of Australia leitet die Arbeit der IIPC Deep Web Archiving Working Group, welche die Identifizierung, Erwerbung, Archivierung, Speicherung und Anzeige von Publikationen und Web-Sites erforscht, die datenbank-gesteuert sind. Diese Kategorie von Web-Ressourcen ist uns ein Anliegen, weil eine wachsende Zahl wichtiger Ressourcen in dieser Form erscheinen. Aus technischen Gründen sind wir zurzeit nicht in der Lage, sie zu sammeln.

Die Ziele der Deep Web Group sind:

- + Identifizierung der von den Verlegern bereitzustellenden begleitenden Informationen, die für abgelieferte benötigt werden,
- + Entwicklung eines Datenmodells und eines Schemas zur Archivierung der Inhalte einer Datenbank in XML und

- + Entwicklung eines Online-Zugriffstools zur Suche und Navigation in den archivierten Datenbanken.

Die Software, die entwickelt wird, wird Open-Domain-Software sein und damit erhältlich für jeden, der daran interessiert ist.

Andere IIPC-Arbeitsgruppen arbeiten an Fragestellungen wie Definition von Anforderungen an die Architektur von Archiven, Identifizierung der Bedürfnisse von Forschern bei der Benutzung von Webarchiven und Entwicklung von Zugriffstools und Webcrawling-Software.

(ii) Internationale Konferenz über die Archivierung von Web-Ressourcen
(<http://www.nla.gov.au/webarchiving/>)

Eine internationale Konferenz über die Archivierung von Web-Ressourcen und andere Probleme der sammelnden Institutionen findet in diesem Jahr vom 9.-11. November in der Nationalbibliothek in Canberra statt. Folgende Veranstaltungen finden im Zusammenhang mit der Konferenz statt:

- + Ein Treffen des IIPC Steering Committee,
- + ein „Australien-Tag“ am 8., an dem hauptsächlich aktuelle australische Projekte im Zusammenhang mit digitalen Ressourcen vorgestellt werden und
- + ein Informationstag am 12., an dem Tools und Methoden zur Web-Archivierung aus allen Teilen der Welt vorgestellt werden.

Das Hauptziel dieser Konferenz ist es, wichtige Problemfelder zu identifizieren, mit denen die Cultural-Heritage-Institutionen bei der Sammlung von Web-Ressourcen konfrontiert sind, und zu untersuchen, wie diese international angegangen werden. Besonderes Augenmerk wird dabei auf eine Überprüfung der Ansätze gerichtet, nach der die sammelnden Institutionen momentan vorgehen und darauf, aktuelle Forschungsprojekte ausfindig zu machen, die die verschiedenen Ansätze unterstützen. Die Konferenz ist nicht technisch ausgerichtet, sondern konzentriert sich vielmehr auf die betrieblichen und strategischen Bedürfnisse der sammelnden Institutionen.

Informationen, die bei der Konferenz zusammengetragen werden und die für ICABS im Zusammenhang mit Web-Archivierung relevant sind, werden genutzt, um die Informationen, die bereits über PADI zugänglich sind, anzureichern.

(iii) Automatisierung der Speicherung und Archivierung von Online-Amtsdruckschriften

Anfang 2003 hat die Bibliothek das Commonwealth Metadata Pilot Project gestartet, das zwei Hauptziele hat:

- + Verbesserung der Ermittlung Aufspürens von Online-Publikationen des Australian Commonwealth Government und des Zugangs dazu mittels der National Bibliographic Database (NBD), die über den Dienst (<http://www.nla.gov.au/kinetica>) zugänglich ist und
- + Rationalisierung der Identifizierung und der Archivierung dieser Publikationen in PANDORA oder einem anderen öffentlichen Archiv zur Sicherstellung des dauerhaften Zugangs über NBD.

Ein grundlegendes Ziel ist die Überprüfung der notwendigen Metadatenstandards und die Entwicklung und Veröffentlichung von Richtlinien, um die teilnehmenden Institutionen zu ermutigen, qualitativ hochwertige Metadaten zu erstellen.

An dem Pilotprojekt, das bis 2005 dauern wird, ist eine Reihe von staatlichen Institutionen beteiligt. Es besteht aus folgenden Phasen:

- + Erforschung der Möglichkeiten, Metadaten von den Partner-Institutionen zu erhalten (einschließlich Harvesting),
- + Konvertierung der gesammelten Metadaten von Dublin Core nach MARC21 mithilfe von MODS, um sie in NBD einzuspielen,
- + Automatisierung von Harvesting, Konvertierung, Einspielung und Aktualisierung von Metadaten sowie
- + Verbesserung der PANDAS-Software, um automatisiertes Harvesting und Speichern durch die Metadaten auszulösen. (Das soll im Laufe des Jahres 2005 geschehen.)

(iv) Erforschung und Förderung von Methoden der Archivierung von Netzpublikationen, die durch Web-Harvesting gesammelt wurden

Nützliche Informationen zu diesem Thema sind bereits über PADI erhältlich und wie bereits erwähnt, wird die Konferenz über Web-Archivierung eine Plattform bieten, um weitere bereits angewendete Verfahren und aktuelle Entwicklungen zu untersuchen. Zusätzlich wird von einer Reihe kultureller Institutionen ein Überblick laufender Aktivitäten erstellt, um sicherzustellen, dass auch diese in PADI vertreten sind.

Um die Informationen mittels PADI leichter zugänglich zu machen, wird PADI im Jahr 2005 umstrukturiert. Die Neustrukturierung wird sich auf einige aktuelle Probleme mit dem System beziehen, die die Recherche von Ressourcen betreffen. Sie zielt darauf ab, PADI durch eine Neuordnung von Schlüssel-Dokumenten nach häufig nachgefragten Themen – einschließlich der digitalen Archivierung - nützlicher zu machen und diese Themen (die dann „Trails“ (Pfade) genannt werden) über die PADI-Homepage zugänglich zu machen.

Langzeiterhaltung digitaler Materialien (ICABS Action 3.4)

Das Ziel unserer Aktivitäten in diesem Arbeitsfeld ist es, in Zusammenarbeit mit der Preservation and Conservation Section der IFLA, eine Übersicht über existierende Standards, Richtlinien und Regelwerke zur Erhaltung digitaler Materialien zu erstellen.

Die Entwicklung von Programmen zur Erhaltung digitaler Materialien ist eine scheinbar unüberwindbare Herausforderung für viele Bibliotheken. Die Absicht, die hinter unseren Aktivitäten auf diesem Gebiet steht, ist es, Faktoren ausfindig zu machen, die die Bereitschaft der Bibliotheken zur Mitwirkung beeinflussen und Formen der Anleitung zu finden, die Bibliotheken auf den Einsatz von Verfahren vorbereiten oder die ihre bereits existierenden Programme effektiver gestalten.

Ein zentraler Bestandteil dieser Vorgehensweise ist die Benutzung von PADI, um bestehende Anleitungen zu finden und um diese leichter zugänglich zu machen. Bedeutende Lücken werden ebenfalls erkannt und bis Mitte 2004 in PADI ergänzt. Wie bereits erwähnt, werden der Aufbau und die Schnittstellen von PADI umstrukturiert, damit Informationen zu Schlüsselthemen wie Standards und Richtlinien leichter zugänglich sind.

Die Beziehungen der Bibliothek zu anderen Bibliotheken in den Regionen Südostasien und Pazifik bieten ebenfalls eine hervorragende Möglichkeit, einzuschätzen, welche Art von Orientierungshilfe in einer großen, verschiedenartigen Region am hilfreichsten wäre. Die Bibliothek dient dem IFLA Preservation and Conservation Program als Regionales Zentrum für Ozeanien und Südostasien und unterhält Arbeitsbeziehungen zu der Kampagne zur Erhaltung des digitalen Erbes der UNESCO Information Society Division.

Im November 2002 hat die UNESCO eine regionale Beratung über digitale Erhaltung durchgeführt, in Verbindung mit der Vorbereitung von Richtlinien, mit deren Erstellung die National Library of Australia beauftragt war. Auf der Grundlage der gesammelten Informationen haben wir im Februar 2004 eine Checkliste von Befähigungsfaktoren und eine Anleitung zur Verbesserung der Aufgabenwahrnehmung für einen UNESCO Memory of the World Nomination Workshop in Manila, Philippinen, erstellt. Die Befähigungsfaktoren umfassen beispielsweise das Bestehen einer Sammlung digitaler Ressourcen, Zugang zu Wissen und Fähigkeiten auf dem Gebiet der Informationstechnologie und geeignete Personalressourcen zur Verwaltung der digitalen Sammlung.

Diese Dokumente werden nach dem Feedback aus dem Workshop überarbeitet und im Juni 2004 für eine weitere Verbreitung fertig gestellt. Natürlich werden sie über PADI zugänglich sein.

Übersetzt von Susanne Oehlschläger
Die Deutsche Bibliothek, ICABS-Sekretariat