

## Abstracts

**Stephen Abrams - Harvard University Library (HUL), USA**

### Digital Formats and Preservation

Effective digital preservation requires the proper characterization of the objects under managed care. In order to provide full access to the underlying information content of digital objects, the fundamental characterization property is format. Without proper format typing all digital content is opaque. Format typing permits the interpretation, rendering, and processing of encoded digital content necessary to fulfill the preservation mandate to provide access to stored digital content over time. Major format dependencies can be identified in many of the workflows defined by the OAIS repository reference model. Fortunately, a new generation of format-aware tools and evolving sets of best practices are now becoming available for use by preservation managers. However, in many instances the range of formats that a digital repository is expected to support will exceed the local format expertise of that repository's preservation staff. This fact, coupled with the long time-scales over which an understanding of formats must be maintained, underscores the importance of establishing sustainable public registries of format representation information. This presentation will review the format dependencies found in preservation workflows, survey recent developments in the area of format-aware tools, best practices, and registries, and will conclude with two cases studies of preservation-driven format migrations.

Die effektive Langzeitarchivierung digitaler Objekte erfordert die präzise Beschreibung der zu archivierenden Bestände. Für einen umfassenden Zugriff auf die in einem digitalen Objekt enthaltenen Informationen ist das digitale Format die grundlegende Eigenschaft. Ohne eine präzise Typisierung des Formats wird der digitale Inhalt unsichtbar. Der Formattyp erlaubt die Interpretation, Darstellung und Bearbeitung des digital kodierten Inhalts, die notwendig sind, um die Aufgabe zu erfüllen, den Zugriff auf gespeicherte digitale Inhalte langfristig zu erhalten. Man kann in vielen der vom OAIS-Referenzmodell definierten Arbeitsabläufen wichtige Abhängigkeiten von den Formaten feststellen. Erfreulicherweise wird den Managern der digitalen Archive mittlerweile eine neue Generation von Werkzeugen und Verfahrensweisen zur Verfügung gestellt, die die zugrundeliegenden Formate berücksichtigen. In vielen Fällen übersteigt jedoch die Vielzahl der von einem Archiv zu unterstützenden Formate die Kompetenz der Mitarbeiter des Archivs. Diese Tatsache, in Verbindung mit dem weiten Zeithorizont, der für die Erhaltung und Interpretation der Formate gesetzt wird, unterstreicht die Bedeutung der Einrichtung öffentlicher Registrierungsmechanismen für die digitalen Formattypen. Diese Präsentation wird einen Überblick über die im Archivierungsablauf gegebenen Formatabhängigkeiten bieten, neue Entwicklungen im Bereich der Werkzeuge und Verfahrensweisen, die die Formate berücksichtigen, darstellen und mit zwei Fallstudien enden, in denen eine Format-Migration aus Archivierungsgründen beschrieben wird.

**Reinhard Altenhöner - Die Deutsche Bibliothek (DDB), Germany**

### Information Lifecycle Management and long term preservation - technical aspects of workflow organization

Publications were traditionally distributed in the form of print media, but with the advent of global electronic networks many are now being disseminated as digital documents. This evolutionary shift significantly changes the task profile of libraries and especially concerns archive libraries, which are entrusted with the preservation of a nation's cultural heritage. As libraries expand their scope to accommodate digital objects, they can no longer resort to established methods for archiving print, but must adapt their organisation and operation to meet a completely new set of requirements.

These new requirements are spelled out within the framework of Information Lifecycle Management (ILM), which is used in industry and commerce as a standard model for such analysis. This analysis also enables Die Deutsche Bibliothek's present situation to be mapped out in a generalized manner. ILM boils down to a strategic model for actively managing the storage and retrieval of digital objects throughout their entire life cycle.

Which guidelines, methods and instruments are best suited to provide a cost-effective and mission-adequate infrastructure for archived data and their accessory applications?

Die Deutsche Bibliothek is currently leading the cooperative development of a deposit system for digital resources (kopal: Co-operative Development of a Long-Term Digital Information Archive), which will be presented as a background for discussing key aspects of these requirements and their consequences for systems implementation. A final step will illustrate the practical impact of these consequences on the operating processes which are currently being developed.

Veröffentlichungen werden traditionsgemäß in gedruckter Form verbreitet, aber mit dem Aufkommen des Internets werden heutzutage viele Dokumente in digitaler Form zur Verfügung gestellt. Diese evolutionäre Entwicklung verändert das Aufgabenprofil der Bibliotheken entscheidend. Dies betrifft insbesondere Bibliotheken mit bestimmten Archivierungsfunktionen, die für die Erhaltung des nationalen kulturellen Erbes verantwortlich sind. Da Bibliotheken ihr Aufgabenfeld auf digitale Objekte ausweiten, können sie sich nicht nur auf das Archivieren gedruckter Dokumente beschränken, sondern müssen ihre Organisationsstruktur und ihre Arbeitsabläufe den neuen Herausforderungen anpassen.

Diese neuen Anforderungen sind im Rahmen des "Information Lifecycle Management (ILM)" formuliert, das sowohl in der Wirtschaft als auch in der Industrie als ein Standardmodell für die Analyse dieser Anforderungen genutzt wird. Die Analysen ermöglichen es auch der Deutschen Bibliothek ihre aktuelle Situation abzubilden. ILM fungiert hier als ein strategisches Modell, das die Aufbewahrung und das Retrieval der digitalen Objekte über deren gesamte Lebensdauer managt.

Doch welche Richtlinien, Methoden und Instrumente sind für die archivierten Daten und ihre zugehörigen Anwendungen geeignet, die kosteneffektiv sind und eine Infrastruktur haben, die den Aufgaben angemessen ist?

**Gerard Clifton - National Library of Australia (NLA), Australia**

### Safe havens on a choppy sea: digital object management workflows at the National Library of Australia.

The National Library of Australia collects, archives and provides access to a wide range of digital materials, including items from its image, map, manuscript and audio digitisation programmes, as well as significant Australian online resources harvested from the Web. The Library seeks to provide a safe place for the storage, management, preservation and delivery of its digital collections and has developed a layered digital services architecture to support these aims.

This paper provides an overview of the Library's digital services architecture and the systems used to manage its main digital collections. It describes the design of the "Digital Collections Manager" and details the ingest and management workflows for archiving image and sound materials, including assignment of persistent identifiers and recording of object relationships, process histories and technical metadata. An overview of PANDAS, the Library's Web archiving and management system for PANDORA, Australia's Web Archive, is also provided.

Die Nationalbibliothek von Australien sammelt, archiviert und bietet Zugang zu einer großen Auswahl an digitalem Material. Dies beinhaltet vielfältige digitalisierte Objekte (Bilder, Karten, Manuskripte und Audiomaterial) sowie signifikante australische Online-Ressourcen, die mittels Harvesting-Technologie aus dem Internet gesammelt werden. Die Bibliothek ist bestrebt, einen sicheren Ort für die Aufbewahrung, das Management, die Langzeitarchivierung und die Bereitstellung ihrer digitalen Sammlungen zu bieten und hat daher eine vielschichtige digitale Servicearchitektur entwickelt, um diesem Anspruch gerecht zu werden.

Dieser Vortrag bietet einen Überblick über diese digitale Servicearchitektur der Bibliothek und stellt die Systeme und Programme vor, die zum Management der großen digitalen Sammlungen genutzt werden. Die Präsentation erläutert das System/Programm "Digital Collections Manager" und beschreibt detailliert die einzelnen Arbeitsschritte hinsichtlich der Einspeisung und Verwaltung von archiviertem Bild- und Tonmaterial. Dazu gehört auch die Zuweisung von "Persistent Identifiers", die Abbildung von Beziehungen zwischen den Objekten, die Dokumentation der Historie von Prozessen und der technischen Metadaten. Abschließend wird PANDAS vorgestellt; ein System der Nationalbibliothek zur Archivierung und zum Management von Web-Ressourcen für PANDORA, das Australische Web-Archiv.

**Robin Dale - Research Library Group (RLG), USA**

### PREMIS: Implementation & preservation metadata

PREservation Metadata: Implementation Strategies working group, or PREMIS, has been an activity jointly sponsored by OCLC and RLG, focusing on

issues associated with implementing preservation metadata in digital archiving systems. It was established to develop a common, implementable core set of metadata elements for digital preservation, since most published specifications for preservation-related metadata are either implementation-specific or broadly theoretical. Comprised of nearly 30 international experts representing national and university libraries, museums, archives, government agencies, and the private sector, PREMIS was charged to define a set of semantic units that are implementation-independent, practically-oriented, and likely to be needed by most preservation repositories as well as to identify and evaluate alternative strategies for encoding, storing, managing, and exchanging the core elements within a digital archiving system.

As part of the PREMIS work, existing preservation repositories were surveyed about their architectural models and metadata practices. The result, documented in the report *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community*, summarized survey responses which addressed the mission, policy, economic, and technical aspects of digital repositories, as well as current practices for creating, managing, and maintaining preservation metadata within the repository environment. Analysis of the survey responses suggests that the digital preservation community is beginning to coalesce around several emerging trends and best practices in the use and management of preservation metadata.

In May 2005, the PREMIS working group released its final deliverable: the final report. The semantic units that support long term preservation are represented in a data dictionary with implementation details and in METS-compatible XML schemas. Examples of the metadata elements applied against a range of digital objects are also a component of the report. A glossary of terms and concepts, a data model, and a typology of relationships developed during the course of the work further supplements PREMIS's contributions to the cultural heritage community.

The presentation will discuss the work of PREMIS, the resulting reports and preservation metadata tools, as well as the ongoing PREMIS maintenance activities.

Die Arbeitsgruppe PREMIS (PREservation Metadata: Implementation Strategies) wurde gemeinsam von OCLC und RLG ins Leben gerufen und befasste sich mit Metadaten zur Langzeitarchivierung in digitalen Archivsystemen. Sie wurde eingerichtet, um einen gemeinsamen einsatzfähigen Kern von Metadaten Elementen für die Erhaltung digitaler Bestände zu entwickeln, da die meisten öffentlichen Spezifikationen für Metadaten im Zusammenhang mit Archivierung entweder nur für spezifische Zwecke gedacht oder allgemein theoretisch sind. Die Gruppe bestand aus fast 30 internationalen Experten, die National- und Universitätsbibliotheken, Museen, Archive, staatliche Einrichtungen und den privaten Sektor repräsentierten, und wurde damit beauftragt, einen Satz von Begriffen zu definieren, die systemunabhängig und praxisorientiert sind und von den meisten Archiven wahrscheinlich gebraucht werden können. Außerdem sollte sie alternative Strategien für die Kodierung, die Speicherung, die Verwaltung und den Austausch der Kernelemente in einem digitalen Archiv finden und bewerten. Als Teil der Arbeit von PREMIS wurden bestehende Archive über ihren Aufbau und ihre Metadatenpraxis befragt. Das Ergebnis, dokumentiert im Report *"Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community"*, fasst die Ergebnisse zusammen, die die Aufgabe, die Regeln, ökonomische und technische Aspekte digitaler Bestände sowie die bestehenden Verfahren für die Erstellung, Verwaltung und Erhaltung von Archivierungs-Metadaten innerhalb der Bestände betreffen. Die Analyse der Antworten legt nahe, dass die Fachwelt um die digitalen Archive herum beginnt, sich um entstehende Trends und Verfahrensweisen in der Nutzung und Verwaltung von Archivierungs-Metadaten zu konzentrieren.

Im Mai 2005 veröffentlichte die PREMIS-Arbeitsgruppe ihr endgültiges Resultat: den Abschlussbericht. Die Begriffe, die die langfristige Bewahrung digitaler Bestände unterstützen, werden in einem Wörterbuch mit Details über ihren Einsatz und in einem METS-kompatiblen XML Schema präsentiert. Beispiele für die Datenelemente, angewandt auf verschiedene digitale Objekte, sind ebenfalls im Bericht enthalten. Ein Glossar von Begriffen und Konzepten, ein Datenmodell und eine Typologie von Beziehungen, die im Verlauf der Arbeit entwickelt wurden, ergänzen den Beitrag von PREMIS zur digitalen Erhaltung des kulturellen Erbes.

Die Präsentation erörtert die Arbeit von PREMIS, die daraus resultierenden Berichte und Werkzeuge für Archivierungs-Metadaten sowie die laufenden Aktivitäten zur Pflege von PREMIS.

**Michael Day - UKOLN, United Kingdom**

#### **Towards a framework for integrating digital preservation research**

In the decade since publication of the report of the Task Force on "Archiving of Digital Information", there has been a plethora of research initiatives focused on the challenges that surround the long-term preservation of digital content. These initiatives have collectively been very successful in raising awareness of the digital preservation challenge, and have also produced a wide range of output types, e.g., policy documents, reference models, training materials, standards, software, etc.

However, the realities of project organisation and funding mean that digital preservation research is usually specific to a particular organisation or professional world-view. The current challenge is to integrate these fragmented initiatives into a coherent multidisciplinary research framework.

The presentation will provide an outline of how this challenge is now being addressed, focusing on the work of the digital preservation cluster of the DELOS Network of Excellence on Digital Libraries, the UK Digital Curation Centre, and other initiatives.

In den zehn Jahren nach der Veröffentlichung des Berichts der Projektgruppe "Task Force on Archiving of Digital Information" entstand eine Vielzahl an Forschungsinitiativen, deren Fokus sich auf die Herausforderungen richtete, die der Langzeitarchivierung von digitalen Objekten immanent sind. Diese Initiativen waren sehr erfolgreich in ihrem Bestreben, das Bewusstsein für diese Herausforderungen der digitalen Langzeitarchivierung zu schärfen. Darüber hinaus entwickelten sie eine Reihe vielfältiger Ergebnisse wie z.B. Richtlinien, Referenzmodelle, Schulungsmaterialien, Standards, Software etc.

Dennoch zeigt die Realität der Projektorganisation und -finanzierung, dass die Forschung im Bereich der digitalen Langzeitarchivierung auf eine bestimmte Organisation oder eine bestimmte fachliche Weltanschauungen ausgerichtet ist. Die derzeitige Herausforderung besteht daher darin, die einzelnen Initiativen in einem kohärenten interdisziplinären Forschungsrahmen zusammenzuführen.

Dieser Vortrag skizziert, wie diese Herausforderung derzeit angegangen wird; der Fokus liegt hierbei auf Arbeitsergebnissen des DELOS Kompetenzzentrums für digitale Bibliotheken, des britischen "Digital Curation Centre" und anderer Initiativen.

**Markus Enders - Göttingen State and University Library (SUB), Germany**

#### **METS: from presentation to preservation**

As a result of different solutions to describe documents, which had been invented with digitization efforts in US libraries in the mid 90ies, the "Metadata Encoding and Transmission Standard" derived. The idea was to use a common format for different media types (text, image, audio and video streams), which is very flexible to support different document models and allows the usage of several metadata formats.

The "Metadata Encoding and Transmission Standard" (METS) is a container format to describe digital documents. Derived from early metadata description formats as used in "Making of America", it provides a flexible container format not only for digitized printed material but also for other digital media as audio and video streams. A METS-file keeps all information about the document's associated objects together - including structural data, metadata and contentfiles.

Though METS allows the description of complex objects, METS files can be very simple. The only required object is a structural map which may provide optional links to metadata- and fileobjects.

The first METS files were used for presentation purposes of digitized material using XSLTs to create a TOC from the structural map and to implement simple page turner-software to browse image by image through the book (METS file contains metadata about the images and provides a link to the TIFF or JPEG file). As documents are getting more complex and the requirements to store the content are changing, more and more institutions are using document management systems (DMS) to implement repositories. A DMS provides additional and improved functionalities for retrieval, management of content files, template engines, administration toolkits etc. This functionality becomes available for METS files as well, as more and more DMS support METS as their Submission Information Package (SIP).

Preserving content for a long time is still a technical challenge, which requires specialized systems. These systems are not only responsible to preserve the bitstreams, but also for storing all necessary administrative data, which might be useful for future migration etc. These systems are usually compliant with the OAIS model using SIPs and DIPs for ingesting and exporting data from the archive. METS as a standardized container format is a good choice to import/export data to/from OAIS compliant systems. Preservations metadata as defined by PREMIS fits into the administrative metadata section of a METS document.

Currently the preservation requirements are driving the ongoing development of METS. For version 1.5 of the METS schema an extension is discussed to allow pointing into file-archives as tar and zip to e.g. attach metadata (and other associated objects) to embedded files.

Als ein Ergebnis verschiedener Lösungen für die Beschreibung von Dokumenten, die im Rahmen von Digitalisierungsprojekten in den USA Mitte der neunziger Jahre entstanden sind, ist der "Metadata Encoding and Transmission Standard (METS)" entwickelt worden. Die Idee war, ein gemeinsames Format für verschiedene Medientypen (Text, Bild, Audio und Video) zu nutzen, das sehr flexibel verschiedene Dokument-Modelle

unterstützt und die Nutzung verschiedener Metadatenformate erlaubt.

Der "Metadata Encoding and Transmission Standard" (METS) ist ein Containerformat zur Beschreibung digitaler Dokumente. Basierend auf frühen Metadaten-Beschreibungsformaten, wie sie beispielsweise bei "Making of America" genutzt wurden, eignet sich dieses flexible Containerformat nicht nur für digitalisierte Druckwerke, sondern auch für weitere digitale Medien, wie Audio oder Videomaterial. Eine METS-Datei besteht aus dem digitalen Objekt ("content file") und weiteren mit dem Objekt verknüpften Informationen, wie den Struktur-Metadaten oder den deskriptiven Metadaten.

Obwohl METS die Beschreibung komplexer Objekte erlaubt, können METS-Dateien sehr einfach sein. Die einzige Voraussetzung ist eine sog. "structural map", eine Abbildung der Struktur der METS-Datei, welche optionale Verknüpfungen zu Metadaten und Dateien bereitstellt.

Zunächst wurde METS zu Präsentationszwecken eingesetzt: XSLTs wurden genutzt, um ein Inhaltsverzeichnis aus der "structural map" zu generieren oder um einfache Software einzusetzen, die Bild für Bild durch ein digitalisiertes Buch blättert (die METS-Datei enthält die Metadaten über das Image und stellt den Link zu den TIFF oder den JPEG Dateien bereit). Da digitale Dokumente bzw. die Dokumentstrukturen komplexer werden und sich auch die Anforderungen an die Speicherung verändern, verwenden mehr und mehr Institutionen sog. Dokumenten-Management-Systeme (DMS), um Repositories zu verwalten. Ein DMS verfügt über zusätzliche und verbesserte Funktionalitäten zum Retrieval, zum Management der inhaltlichen Dateien, zur Generierung von Vorlagen oder auch zur Administration. Auch METS-Dateien können zunehmend diese Funktionalitäten nutzen, da mehr und mehr Dokumenten-Management-Systeme METS als Format für die Dateneinspeisung akzeptieren.

Inhalte über einen längeren Zeitraum zu archivieren, stellt immer noch eine große technische Herausforderung dar, die spezialisierte Systeme erfordert. Diese Systeme müssen nicht nur den Bitstream archivieren, sondern auch alle administrativen Informationen bzw. Daten, die beispielsweise für künftige Migrationen oder ähnliches benötigt werden. Diese Systeme sind in der Regel konform zum Open Archival Information System (OAIS) Modell und nutzen Submission Information Packages (SIPs) und Dissemination Information Packages (DIPs), um Daten in das Archiv aufzunehmen oder wieder auszugeben.

METS als standardisiertes Containerformat ist eine gute Wahl, um Daten in oder aus einem OAIS System zu importieren bzw. zu exportieren.

Metadaten zur Langzeitarchivierung, wie sie von der PREMIS Arbeitsgruppe definiert wurden, können im Abschnitt für administrative Metadaten eines METS-Dokumentes gespeichert werden.

Zur Zeit bestimmen die Bedürfnisse der digitalen Langzeitarchivierung die Weiterentwicklung von METS. Für die Version 1.5 des METS-Schemas wird eine Erweiterung diskutiert, die es erlaubt, in komprimierte Datenarchive wie ZIP oder TAR zu verlinken.

## **Manjula Patel - UKOLN, United Kingdom**

### **A National Preservation Policy for the UK - Thoughts and Challenges**

This talk will describe the state of the art of digital curation in the UK. A number of ongoing projects will be touched on. The activities of the recently formed Digital Curation Centre will be discussed in some detail. In addition some of the national legislative and funding drivers will be outlined.

## **John Kunze - California Digital Library (CDL), USA**

### **Future-Proofing the Web: What We Can Do Today**

All we know about predicting our digital future is based on our past, a review of which reveals a remarkable truth dating from the beginning of the digital era: plain text is a versatile and lossless format that is just as readable with today's computers as it was 30 years ago. Compared to the fonts, colors, point sizes, and graphics available in contemporary formats, plain text may look dull and dry, but in fact this "desiccated data" successfully represents all the protocols that built the Internet.

Moreover, it is hard to imagine its not being nominated as the most likely current format to be readable 30 years from now.

A proposed strategy for preserving today's web formats is to go ahead and save the original format, but also to automatically derive and save various "desiccated" versions that, while failing to capture all the original format's richness, nonetheless capture its essential nutrient value. In the case of a document, saving a plain text format version alongside the original would provide a fall back in case the original format failed. One never knows if there will ever be money enough to touch a preserved object again, let alone migrate its format. The effort and storage for a derived plain text version is often needed anyway to support search indexing.

Generalizing, the lesson appears to be that the simpler technological intermediation required to render the digital object for the user, the easier it is to reproduce that intermediation, hence to carry forward the object. Along these lines, the image format analog of plain text file might be a basic raster file, in which the array of pixels (picture elements) could be seen to mimic ancient weaving technology. It may be that adding the complication of a simple run-length encoding compression would be worth the space savings. A strategy for deriving and saving raster images of original documents rendered with today's software has two advantages: we will never have better rendering tools for today's formats than today (with all the features and error-compensation that make malformed format instances -- very common -- renderable) and it provides an additional fall back in case the original and the plain text fail. This is something we can do for preservation today that we may never have the money or the knowledge to do in the future.

Unser heutiges Wissen über die digitale Zukunft basiert auf der Vergangenheit; betrachtet man diese Vergangenheit, so enthüllt sich eine bemerkenswerte Wahrheit, die bis in die Anfänge des digitalen Zeitalters zurückreicht: einfacher Text ("plain text") ist ein vielseitiges und verlustfreies Format, das mit heutigen Computern noch genauso lesbar ist wie es vor 30 Jahren war. Im Vergleich zu den Zeichensätzen, Farben, Schriftgrößen und Grafiken, die in den heutigen Formaten zur Verfügung stehen, mag reiner Text langweilig und trocken wirken, tatsächlich stellen diese "desiccated data" erfolgreich alle Protokolle dar, die das Internet aufbauen.

Darüber hinaus ist es schwer vorstellbar, dass dieses Format nicht auch dasjenige ist, das am ehesten in 30 Jahren noch lesbar sein wird.

Eine mögliche Maßnahme für die Erhaltung der heutigen Internetformate ist, die Originalformate zu archivieren aber gleichzeitig automatisch "desiccated" Versionen zu erzeugen und zu speichern, die zwar nicht die Fülle des Originalformats, wohl aber seinen essentiellen Nährwert erfassen können. Im Falle eines Dokuments würde die parallele Speicherung eines "plain text" Formats neben dem Originalformat eine Sicherung bieten, falls das Originalformat nicht mehr lesbar ist. Niemand weiß heutzutage, ob jemals genug Geld zur Verfügung stehen wird, um die einmal archivierten Objekte je wieder anzufassen oder sie gar in ein neues Format zu migrieren. Der Aufwand für die Erstellung und der Speicherplatz für die "plain text" Version wird oft sowieso benötigt, um den Aufbau eines Suchindex zu unterstützen.

Im allgemeinen scheint die Lehre zu sein, dass, je einfacher die benötigte technologische Vermittlung zur Präsentation eines digitalen Objektes für die NutzerInnen ist, desto einfacher ist es, diese Vermittlung zu reproduzieren und damit das Objekt nutzbar zu halten. Dieser Argumentation folgend, könnte das Bildformat, analog zu dem einfachen Text, eine Rasterdatei sein, in der die Anordnung der Pixel (Bildelemente) alte Webtechniken imitiert. Möglicherweise würde sich der Aufwand lohnen, dadurch Platz zu sparen, indem eine einfache Komprimierung durch Kodierung gleicher Werte durchgeführt wird. Eine Strategie zur Erzeugung und Speicherung von Rasterbildern aus den Bildern, die mit der heutigen Software aus den Originaldokumenten dargestellt werden, bietet zwei Vorteile: Wir werden nie bessere Darstellungswerkzeuge für die heutigen Formate besitzen (mit all den Einstellungen und Fehlerkorrekturen, die die Darstellung fehlerhafter Objekte, wie sie häufig vorkommen, ermöglichen) und es bietet eine zusätzliche Sicherung, falls das Original und die "plain text" Version nicht mehr darstellbar sind. Dies sind Maßnahmen für die digitale Langzeitarchivierung, die wir heute durchführen können, für die wir in der Zukunft jedoch vielleicht nie das Geld oder das Wissen haben werden.

## **Chunming Li - National Library of China (NLC), China**

### **Digital Preservation Activities in the National Library of China**

The Internet, as the fourth information medium, has gradually been accepted by the Chinese people. With Web information increases dramatically day by day, the National Library of China, as a deposit library, began its study on how to preserve Chinese Web information in 2003.

This presentation mainly introduces what the National Library of China has done on:

- a) preservation of Web information, including preservation objects selection, technology solutions, working model, and the working process,
- b) promoting long term digital preservation at the library of domestic publications through cooperation with the large Chinese digital information providers such as Tsinghua Tongfang, Wanfang Data, and Founder Electronics Co.Ltd.

Allmählich wächst bei der chinesischen Bevölkerung die Akzeptanz des Internets als viertem Medium. Aufgrund der täglich ansteigenden Informationsfülle im Internet begann die "National Library of China" in ihrer Eigenschaft als Archivbibliothek im Jahre 2003 damit, sich mit Wegen und Möglichkeiten zur langfristigen Archivierung chinesischer Webinhalte auseinander zu setzen.

Dieser Vortrag stellt die bisherigen Aktivitäten der "National Library of China" auf diesem Gebiet vor:

- a) Langzeitarchivierung von Webinformationen einschließlich Selektionskriterien bzgl. der zu archivierenden Objekte, technischer Lösungen,

Arbeitsmodell und Arbeitsprozesse

b) Förderung von Langzeitarchivierung an der für nationale Publikationen zuständigen Bibliothek durch Kooperation mit den großen chinesischen Anbietern von digitalen Informationen, wie zum Beispiel Tsinghua Tongfang, Wanfang Data und Founder Electronics Co.Ltd.

## **Julien Masanès - European Digital Archives, France**

### **Web Archiving: a collaborative effort in progress**

The web is a virtually infinite information space, and archiving its entirety, all its aspects, is a Utopia. The amount of information presents a challenge, but it is neither the only, nor the most limiting one given the continuous drop in storage device costs. Significant challenges lie in the management and technical issues of the location and collection of web sites. They will be briefly presented.

Because of this, archiving the web is a task that no single institution can fulfil alone. It requires a collaborative effort in which there is a natural division of roles in space (breadth and depth of the information space) as well as a potential functional division of tasks.

But working collaboratively in this domain entails a methodological clarification as well as the definition of common standards & tools in a domain where traditional preservation practices have been unprecedentedly challenged. Progress in this domain will be presented, mainly carried out by the International Internet Preservation Consortium (IIPC) a consortium of 12 national libraries and the Internet Archive.

Web archiving requires also the mastering of complex processes in an extremely diverse and continuously changing technological environment. The emergence of a new type of technology partners for cultural institutions will also be presented with the example of collaboration between the Internet Archive and the European digital Archive with archives and libraries.

Das Internet ist ein unendlicher virtueller Informationsraum; es komplett mit all seinen Aspekten archivieren zu können ist utopisch. Die Menge an Informationen an sich stellt schon eine große Herausforderung dar, jedoch ist sie weder die einzige, noch die größte Herausforderung bezogen auf die stetige Abnahme der Kosten für die Speichermedien. Die Lokalisierung und die Sammlung von Webseiten stellen große Herausforderung an die organisatorische und technische Handhabung. Dies soll hier kurz erläutert werden.

Die Archivierung des Internets kann nicht eine einzige Institution allein bewältigen. Dazu benötigt man kooperative Strukturen, die die natürliche Aufteilung des World Wide Web (Weite und Tiefe des Informationsraumes) und die potenzielle funktionale Aufteilung der Aufgaben regeln. Arbeitsteilige Kooperation auf diesem Gebiet benötigt jedoch eine methodische Klärung sowie die Definition gemeinsamer Standards und Werkzeuge; und dies in einem Bereich, in dem Langzeitarchivierung traditioneller Objekte beispielhaft umgesetzt wurde. Entwicklungen auf diesem Gebiet werden hier vorgestellt, die zum größten Teil von dem "International Internet Preservation Consortium (IIPC)" (ein Konsortium aus zwölf Nationalbibliotheken und dem Internetarchiv) vorangetrieben wurden. Webarchivierung erfordert die Leitung komplexer Prozesse innerhalb eines vielfältigen und sich fortwährend ändernden technischen Umfelds. Die Entstehung eines neuen Typus von technologisch speziell auf diesem Gebiet erfahrenen Partnern der kulturell verantwortlichen Institutionen wird anhand des Beispiels der Kooperation zwischen dem "Internet Archive" und dem "European Digital Archive" mit Archiven und Bibliotheken vorgestellt.

## **Elisabeth Niggemann - Die Deutsche Bibliothek (DDB), Germany**

### **Preserving our digital heritage - weaving the web of trust**

Many digital objects have a lasting value and significance. As an important part of our cultural heritage they need to be collected and preserved for current and future generations of researchers. In the analogue world we have established criteria for selection, standardized technologies and well defined responsibilities for preservation. In the digital world, responsibilities, technology and selection have to be redefined and established. There is an urgent need to do this, because digital objects have a limited shelf life and the risk of loss is therefore imminent. This is especially true for the preservation of the great number of freely floating objects in the World Wide Web. But the mere storage of digital content is not even enough. A true preservation strategy also needs to ensure the long term accessibility to and usage of the stored objects. It is this combination of long term storage and long term usability that defines a "trusted digital repository".

Developing the right technology and establishing the adequate organisational structure is not an easy task, but setting up criteria of selection might even be harder. We need answers to the question what documents researchers need today and what they will need tomorrow or in the far future. We need the right guidelines for selection or valuable resources will have disappeared tomorrow.

Because of the high-speed, high-price, high-risk issues of digital preservation, it can only be addressed in a cooperative way. What is needed is a coordinated web of trusted digital repositories with preservation policies that guarantee long term archiving as much as long term usability.

Viele digitale Objekte besitzen einen langfristigen Wert und auch eine solche Bedeutung. Da sie einen entscheidenden Bestandteil unseres kulturellen Erbes bilden, ist es wichtig, sie für diese und künftige Generationen von Wissenschaftlern zu sammeln und langfristig zu archivieren. Für analoge Objekte gibt es hinsichtlich der Langzeitarchivierung feste Auswahlkriterien, standardisierte Technologien und definierte Verantwortlichkeiten. Für digitale Objekte müssen diese Strukturen jedoch neu definiert und etabliert werden. Es besteht dringender Handlungsbedarf, da das Risiko des Verlusts aufgrund ihrer geringen Lebensdauer außerordentlich hoch ist. Dies trifft besonders auf die Langzeitarchivierung einer Vielzahl von digitalen Objekten zu, die wahllos und unkontrolliert über das World Wide Web verbreitet werden. Die bloße Aufbewahrung dieser digitalen Objekte ist jedoch bei weitem nicht ausreichend. Benötigt werden Strategien zur Langzeitarchivierung digitaler Objekte, die sowohl den Zugang als auch die Nutzung langfristig gewährleisten. Diese Kombination aus Langzeitarchivierung und Langzeitnutzung zeichnet ein zuverlässiges digitales Repository aus.

Es ist keine einfache Aufgabe, die richtige Technologie zu entwickeln und die notwendigen organisatorischen Strukturen bereit zu stellen; die größere Herausforderung ist es jedoch, die Auswahlkriterien zu definieren. Wir benötigen Antworten auf die Frage, welche Dokumente Wissenschaftler heute und in Zukunft benötigen. Wir benötigen ebenso geeignete Richtlinien für die Auswahl dieser Dokumente bzw. Objekte, andernfalls werden wertvolle Ressourcen schon bald verschwunden sein.

Hohe Kosten, rasante Entwicklungen und große Risiken stellen Herausforderungen an die digitale Langzeitarchivierung, die nur gemeinsam und kooperativ bewältigt werden können. Benötigt wird daher ein koordiniertes Netz von vertrauenswürdigen digitalen Repositories und geeignete Strategien zur digitalen Langzeitarchivierung, die langfristig die Erhaltung und Nutzung von digitalen Objekten sicherstellen.

## **Nancy McGovern and Marcy Rosenkrantz - Cornell University Library (CUL), USA**

### **Cornell University Library Open Archival Information System**

Cornell University Library (CUL) has developed nine priority goals for the coming years in support of its strategic plan. CUL's digital preservation efforts are well established, and include a highly acclaimed "Digital Preservation Management" workshop. So it is natural that one of its priorities is to develop an OAIS-compliant digital preservation system for its digital assets within three years.

We will discuss our work plan and progress to date, which includes a "Digital Preservation Policy Framework", a full review of the OAIS reference model that forms our own requirements for such a system, an elevator pitch for the project, and a draft selection policy and associated procedures.

A related project is called "Ensuring Access to Mathematics Over Time", funded by the National Science Foundation. This is a collaborative effort with SUB Göttingen to preserve serial literature in mathematics. SUB Göttingen will preserve journal articles from its digitization of Springer mathematics journals. CUL is preserving mathematics literature from its Project Euclid. In this presentation we will discuss the progress we have made thus far and our plans for the final year of our project. We will discuss the implementation of collaboratively administered, physically distributed digital preservation systems, such as the ones at CUL and SUB Göttingen and the unique set of challenges they present. Scenarios and use cases we have developed, and agreements we've made will be discussed, as well.

Die Cornell University Library (CUL) hat für die kommenden Jahre neun primäre Ziele entwickelt, um ihren strategischen Plan zu verwirklichen. Es gibt bereits Maßnahmen der Bibliothek im Bereich der digitalen Langzeitarchivierung, beispielsweise einen allseits anerkannten "Digital Preservation Management" Workshop. Eine Priorität besteht daher darin, innerhalb der nächsten drei Jahre ein OAIS-kompatibles digitales Langzeitarchivierungssystem für die eigenen digitalen Bestände zu entwickeln.

Wir präsentieren dieses Vorhaben und die bis dato erzielten Fortschritte, welche folgendes umfassen: Rahmenrichtlinien für Strategien zur digitalen Langzeitarchivierung (Digital Preservation Policy Framework), eine ausführliche Darstellung über das OAIS Referenzmodell, das unsere eigenen Anforderungen an ein solches System erfüllt, eine kompakte Selbstpräsentation und den Entwurf eines Dokumentes zu Selektionskriterien und den damit verbundenen Prozessen.

Ein hiermit verwandtes Projekt ist das durch die National Science Foundation geförderte Projekt "Ensuring Access to Mathematics Over Time". Dabei

handelt es sich um ein Gemeinschaftsprojekt mit der SUB Göttingen, bei dem mathematische Serienliteratur langfristig archiviert wird. Die SUB Göttingen archiviert Zeitschriftenartikel der von ihr digitalisierten Springer Mathematikzeitschriften, die Cornell University Library archiviert mathematische Literatur ihres Euclid-Projektes. In unserem Vortrag werden wir den bisherigen Fortschritt auf diesem Gebiet und unsere weiteren Pläne für das letzte Projektjahr erörtern. Unser Vortrag behandelt die Implementierung digitaler Langzeitarchivierungssysteme, wie die der Cornell University Library und der SUB Göttingen, die zwar kooperativ verwaltet werden, sich aber an verteilten Orten befinden, und die damit verbundenen einzigartigen Herausforderungen. Erörtert werden außerdem von uns entwickelte Szenarien und Anwendungsfälle sowie von uns getroffene Übereinkünfte / Abkommen.

## **Stefan Strathmann & Heike Neuroth - Göttingen State and University Library (SUB), Germany**

### **Developing a National Preservation Policy: Experiences in Germany**

One of the versatile tasks of the project nestor is to develop guidelines for the long-term preservation of digital objects in Germany and to propose national preservation strategies.

A national preservation policy forms the framework for the efforts within the long-term preservation and thus determines intentions, aims and responsibilities. A possible preservation policy needs to consider a range of different factors, as e.g.

- the federal structure in Germany and thus the relation between the Federal Government and the several States
- the interests of the particular institutions (archives, libraries, museums, universities etc.)
- the needs which result from managing different material (text, image, audio, primary data etc.) and
- the different professional and financial responsibilities as well as the rights issues that are inherent in the digital preservation.

This paper will summarize the results that have been made in the nestor project concerning the development of a national preservation policy.

Eine der vielfältigen Aufgaben des Projektes nestor besteht darin, Richtlinien für die Langzeitarchivierung digitaler Objekte in Deutschland zu entwickeln und nationale Strategien vorzuschlagen.

Eine nationale LZA-Policy steckt den Rahmen der Bemühungen um die Langzeiterhaltung digitaler Objekte ab und legt Absichten, Ziele und Zuständigkeiten fest.

Eine zu entwickelnde LZA-Policy muss eine Reihe unterschiedlicher Faktoren berücksichtigen wie z.B.

- die föderale Struktur in Deutschland und damit das Verhältnis von Bund und Ländern,
- die Interessen der unterschiedlichen Berufsstände (Archiv, Bibliothek, Museum, Universitäten etc.),
- die Notwendigkeiten, die sich aus der Behandlung unterschiedlicher Materialarten (Text, Bild, Audio, Primärdaten etc.) ergeben und
- Fragen nach den sachlichen und finanziellen Zuständigkeiten sowie juristische Aspekte, die sich aus der digitalen Langzeitarchivierung ableiten.

Der Vortrag wird die bisherigen Ergebnisse zusammenfassen, die im Rahmen des Projekts nestor zur Entwicklung einer nationalen LZA-Policy erzielt wurden.

## **Johan Stapel – National Library of the Netherlands (KB), The Netherlands**

### **Workflow management in an operational archiving environment**

In 2003 the digital archiving system of the National Library of the Netherlands (KB) was taken into production. This system is called the e-Depot and its technical heart is the IBM system called Digital Information Archiving System (DIAS). The KB e-Depot is dedicated to the long-term storage of and access to large quantities of digital publications. Currently, the system has ingested several millions of digital objects. This presentation will explicate how the way we have organized the workflow processes in this operational digital archiving environment.

2003 wurde das digitale Archivsystem der Nationalbibliothek der Niederlande (KB) in Betrieb genommen. Das System heißt e-Depot und ihm liegt als technischer Kern das IBM-System DIAS (Digital Information Archiving System) zugrunde. Das e-Depot der KB dient dazu, große Mengen digitaler Publikationen langfristig zu archivieren und verfügbar zu machen. Zur Zeit sind bereits mehrere Millionen digitale Objekte in das System eingespeist. Dieser Vortrag wird erläutern, wie wir die Workflowprozesse dieses digitalen Archivierungssystems organisiert haben.

## **Xiaolin Zhang and Zhixiong Zhang - Chinese Academy of Sciences (CAS), China**

### **Developing a National Preservation Network for STM e-Journals**

China is facing urgent needs for collaborative digital preservation strategies. The situation is more acute for foreign STM materials critical to key research and educational institutes, because, on one hand, in all of those institutes, STM e-Journals becomes the major part of their library resources and represents the main streams of use, on the other hand, few of them have developed strategies and plans for long-term preservation and fewer have placed these efforts in a nationally coordinated and collaborative context.

In this presentation, following an analysis of current practices in STM digital resources management in China, challenges are briefly explored for implementing an economically sound, service reliable, and organizationally sustainable digital preservation undertaking, and needs are recognized for a nationally collaborative preservation strategy and its corresponding policy framework to guide and coordinate preservation efforts of each and all the institutes.

Structure models for Chinese national collaborative STM preservation networks are elaborated, considering (1) the technological, economic, administrative, and managerial responsibilities incurred by joining the national network, (2) the existing structures and responsibility schemes of current library systems, consortia, and other collaboration arrangements, especially in terms of digital resource development, and (3) various possible responsibility relations among various players, of any national preservation network, such as owners, preserving institutions, repository managers, targeted service audience, and preservation funding agencies, etc. Special consideration is further extended to the requirements for and possible implementation schemes for fail-safe mechanisms and succession planning in different contexts.

Policies for implementing and managing a nationally coordinated distributed network of trustworthy repositories are further discussed, including the selecting and entrusting of repository managers, coordinating and monitoring their preservation ingest, planning and management, and provision of public services, and evaluating operations and management soundness, for openly verifiable and collaboratively manageable repository network; Integration of the national STM digital preservation system into the national science and technology infrastructure (NSTI) is discussed, including incorporating digital preservation into NSTI strategy and policy framework, organizing preservation repositories as part of NSTI services, coordinating preservations of STM publications with preservation efforts of other scientific resources (such as scientific data), and very importantly, channel NSTI funding into development and maintenance of the national STM preservation network.

China steht einem dringenden Bedarf an kooperativen digitalen Erhaltungsstrategien gegenüber. Die Situation ist besonders akut für ausländisches STM-Material, das von grundsätzlicher Bedeutung für Forschungs- und Bildungsinstitute ist. Einerseits machen in all diesen Instituten elektronische STM-Zeitschriften zunehmend den größten und meistgenutzten Teil der Ressourcen aus und andererseits haben nur wenige von ihnen Strategien und Planungen zur langfristigen Bewahrung dieser Bestände entwickelt und noch weniger diese in einen national koordinierten und kooperativen Zusammenhang eingebracht.

Wir beginnen diese Präsentation mit einer Analyse derzeitiger Praktiken in der Verwaltung digitaler Quellen in China, dann werden kurz die Herausforderungen erkundet, die eine ökonomisch stabile, verlässliche und organisatorisch nachhaltige Anstrengung im Bereich digitaler Erhaltung bedeutet. Weiterhin wird der Bedarf an einer nationalen kooperativen Bewahrungsstrategie und einem entsprechenden Rahmenprogramm herausgestellt, welche die Bewahrungsmaßnahmen der einzelnen Institute leitet und koordiniert.

Es werden Strukturmodelle für ein chinesisches nationales kooperatives Bewahrungsnetz für STM ausgearbeitet, unter Berücksichtigung

1. der technologischen, ökonomischen und verwaltungsmäßigen Verantwortung, die eine Teilnahme am nationalen Netz mit sich bringt,
2. der existierenden Strukturen und Verantwortungsbereiche des bestehenden Bibliothekssystems, der Konsortien und anderer kooperativer Einrichtungen, insbesondere in Bezug auf die Entwicklung digitaler Quellen, und
3. verschiedener möglicher Verantwortlichkeiten unter den Teilnehmern an einem nationalen Bewahrungsnetz, wie Besitzern, Archiven, Bestandsmanagern, Zielpublikum, Förderungseinrichtungen usw.

Besondere Berücksichtigung finden schließlich die Erfordernisse an ausfallsichere Mechanismen, mögliche Pläne zu ihrer Implementierung und die Nachfolgeplanung in verschiedenen Zusammenhängen.

Des Weiteren werden Maßnahmen für den Aufbau und die Verwaltung eines nationalen, koordinierten, verteilten Netzes vertrauenswürdiger Archive

diskutiert. Dies schließt die Auswahl und Verantwortung der Bestandsverwalter sowie die Koordinierung und Überwachung der Dokumentaufnahme, Planung und Verwaltung, die Bereitstellung öffentlicher Dienste, die Bewertung der Qualität von Betrieb und Verwaltung für ein öffentlich kontrolliertes und kooperativ verwaltetes Archivierungsnetzwerk ein.

Die Integration des nationalen Erhaltungssystems für digitale STM-Literatur in die nationale Infrastruktur für Wissenschaft und Technologie (NSTI) wird erörtert, einschließlich der Überführung der digitalen Bestandserhaltung in die NSTI-Strategie und –Rahmenprogramme, die Organisation von digitalen Archiven als NSTI-Dienstleistung, die Koordination der Bewahrung von STM-Publikationen mit den Erhaltungsaktivitäten anderer wissenschaftlicher Quellen (wie wissenschaftlicher Daten), und – besonders wichtig – die Ausrichtung der NSTI-Förderung auf die Entwicklung und Erhaltung eines nationalen Netzes zur STM-Erhaltung.