

lands of the United States
THE CHESAPEAKE PROJECT

Legal Information Archive



The Chesapeake Project:

One Model for Digital Preservation

101st AALL Annual Meeting & Conference
Portland, Oregon

July 13, 2008

Agenda

- Introduction
- About The Chesapeake Project
- Project Status (report on First-Year Evaluation)
- Looking Ahead

A History of Lost Legal Information

- Legal information has always been at risk for loss & destruction (whether by natural or human means)
 - 1814: first 3,000 volumes of the Library of Congress destroyed
 - 1966: thousands of books, manuscripts, and legal materials at the Biblioteca Nazionale & Archivio di Stato destroyed in the Arno River flood
 - 2003: The National Library of Iraq burned & looted
 - 2005: Government records & documents destroyed in the flood of New Orleans
 - ALSO...mid-1800s onward: mass-printing processes using acidic paper

Today, A New Threat...

- Important legal materials are increasingly being **digitally born**
- Unprecedented mass acceptance/adoption of digital culture
- 2007: **281 million gigabytes** of digital information created, captured, or replicated — surpassing global storage capacity
- The current “digital universe” will increase 10 times over by 2011

New Risks in the Digital Age

- Archival preservation standard = film; there is no digital archival preservation standard
- ALA adopted its first, accepted definition of digital preservation only last year (June 2007)
 - Short version: "Digital preservation combines policies, strategies and actions that ensure access to digital content over time"
- Digital formats vulnerable due to:
 - Technology obsolescence
 - Unknown life span of digital media
 - CDs = 5 – 59 years
 - digital formats = obsolete in 5 – 20 years

New Risks in the Digital Age

- Legal information permanently lost from the Web as files are removed and URLs changed/ inactivated through routine Web site maintenance activities
- Average lifespan of a Web page = 44 – 75 days

Preserving Digitally Born Legal Information

- This challenge is not insurmountable
- Digital preservation strategies:
 - **Technology Preservation** — Imagine a computer museum....
 - **Migration** — Continual copying of information from old formats to new formats
 - **Emulation** — Designing new applications to mimic old applications
- Current philosophy: shift in focus from preservation action (migration, emulation) to preservation metadata, standards, and architecture

Preserving Digitally Born Legal Information

- Digital repositories and Web harvesters are available for a fee or independent development using open-source software
- Requires commitment of staffing, infrastructure, and funding
 - The long-term viability of a digital archiving program depends on organizational “fitness” as well as technological infrastructure

Law Libraries & Digital Preservation

- Recent survey of academic/state law library directors (37 respondents)
 - Only 4 of 59 reported digital preservation projects involved the preservation of digitally born materials
 - Remaining 55 = preservation of digitized print materials
- Yet, by a margin of 2 to 1, respondents indicated that they believed digitally born materials were in more urgent need of preservation than print materials

Law Libraries & Digital Preservation

- Top factors limiting digital preservation activities:
 - Lack of funding
 - Staffing shortages
 - Lack of staff with requisite technological or digital preservation expertise
- Can be addressed through collaboration = sharing of resources, reducing the burden placed upon a single institution in implementing a digital preservation program

About The Chesapeake Project

- Affiliated with the Legal Information Preservation Alliance, whose membership includes the American Association of Law Libraries
 - Origin = 2003 conference, Preserving Legal Information for the 21st Century
- The Chesapeake Project is a shared Legal Information Archive
 - A collaborative digital preservation program for digitally born materials published online

About The Chesapeake Project

- A two-year (2007-2009) pilot implemented by three law libraries:
 - Georgetown Law Library
 - Maryland State Law Library
 - Virginia State Law Library
- **Project Mission:** to successfully develop and implement a pilot program to stabilize, preserve, and ensure permanent access to critical born-digital legal materials on the World Wide Web.

The Chesapeake Project is working to establish the beginnings of a strong regional digital archive collection of U.S. legal materials as well as a sound set of standards, policies, and best practices that could potentially serve to guide the future realization of a nationwide preservation program.

Selection & Collection Scope

- Individual selection parameters set to guide/limit scope:
 - Maryland and Virginia State Law Library collections = state-issued materials, as well as some community- and organization-published reports and studies
 - Georgetown Law Library collections = thematic, secondary legal materials based on scholarly areas of interest and legal research institutes
- Focus on the capture and preservation of discrete online publications, as opposed to entire Web sites

Digital Preservation Strategies & Tools

- OCLC Digital Archive:
 - Compliant with ISO reference model for an Open Archival Information System (OAIS)
 - OCLC = stable; archive is backed by a sound organizational structure
 - Fair shared pricing structure
 - Bit-level preservation = files remain uncorrupted and renderable in their original formats

Digital Preservation Strategies & Tools

- OCLC responsibilities:
 - Secure onsite storage at OCLC facilities
 - Multiple copies of backup data/disaster tapes at offsite facility
 - Regular schedule of virus-checking, file format verification, and fixity-checking using checksum algorithms

Digital Preservation Strategies & Tools

- Like all things in the digital environment, digital archive systems must evolve to take full advantage of technological progress and evade obsolescence...
- Spring/Summer 2008 = transition from the original OCLC Digital Archive to a more sophisticated, two-tiered digital-preservation and access system
 - Access copy in CONTENTdm + archival masters in dark Digital Archive (similar to original Digital Archive)
 - Added point of access through CONTENTdm interface
 - PDF full-text searchability

Discovery & Access of Archived Items

- Discovery and access made available through:
 - Participating institutions' local OPACs
 - The open-access WorldCat.org system
 - Subscription OCLC FirstSearch/WorldCat databases
 - The Chesapeake Project's new CONTENTdm system
- Bibliographic treatment is crucial to access
 - MARC record created in OCLC prior to harvest
 - Unique digital archive URL hyperlinked to the archived access copy (856 field)
 - MARC metadata crosswalked to Qualified Dublin Core in CONTENTdm



GULLiver

Library Catalog at GEORGETOWN LAW



Start Over



Request



Save Record



MARC Display



Another Search

RECORD #

b491250

Search

Call # KF2994 .S43 2008 electronic

Title The Section 108 Study Group report [electronic resource] : an independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress

Imprint [Washington, D.C. : The Section 108 Study Group], 2008

Click on the following to:

[Connect to electronic version](#)

[Connect to electronic version \(executive summary only\)](#)

[Connect to electronic version in Legal Information Archive](#)

BOOKMARK

Copy Status

Contents

Find Similar Items

Full Record

Access via local OPAC

THE CHESAPEAKE PROJECT

Legal Information Archive

About The Chesapeake Project



Home

Search

Create lists, bibliographies and reviews: [Sign in](#) or [create](#)

Search for items:

Search

[Advanced Search](#)

The Section 108 Study Group report an independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress.



☆☆☆☆☆
(not yet rated)

Type: Internet Resource; English

Publisher: [Washington, D.C. : The Section 108 Study Group], 2008.

OCLC: 221658479

Related Subjects: [Copyright -- United States.](#) | [Copyright -- Electronic information resources -- United States.](#) | [Fair use \(Copyright\) -- United States.](#) | [More Subjects ...](#)

Citations: [Cite this Item](#) | [Export to EndNote](#) | [Export to RefWorks](#)

Additional Info: [Connect to electronic version](#) | [Connect to electronic version \(executive summary only\)](#) | [Connect to electronic version in Legal Information Archive](#)

Access via WorldCat.org

THE CHESAPEAKE PROJECT

Legal Information Archive

About The Chesapeake Project



GEORGETOWN UNIVERSITY LAW LIBRARY



WorldCat Detailed Record

[Research Help](#) | [Live Help](#)

◆ Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.

[Home](#)

[Databases](#)

[Searching](#)

[Results](#)

[Staff View](#)

[My Account](#)

[Options](#)

[Comments](#)

[Exit](#)

[Hide tips](#)

[List of Records](#)

[Detailed Record](#)

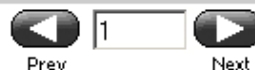
[Marked Records](#)

[Saved Records](#)

Go to page



WorldCat results for: no: 221658479. Record 1 of 1.



Mark:

[Detailed Record](#)

[Add/View Comments](#)



The Section 108 Study Group report

an independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress.

2008

English Internet Resource Computer File 187 p. : digital, PDF file.

[Washington, D.C. : The Section 108 Study Group].

GET THIS ITEM

Access: <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>

Availability: **FirstSearch indicates your institution owns the item.**

- ◆ [Libraries worldwide that own item:](#) 6 **Georgetown University Law Library**
- ◆ [Search GULLiver: the Law Library Catalog](#)

Access via subscription OCLC WorldCat database

THE CHESAPEAKE PROJECT

Legal Information Archive

| About The Chesapeake Project |

THE CHESAPEAKE PROJECT

Legal Information Archive

 search

[home](#) : [browse](#) : [advanced search](#) : [preferences](#) : [my favorites](#) : [collection directory](#) : [about](#) : [help](#)

[login](#)

[add to favorites](#) : [reference url](#)

[back to results](#) : [previous](#) : [next](#)

The Section 108 Study Group report an independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress.

[Open PDF in new window](#) | [Go to PDF description](#)

The screenshot shows a PDF viewer interface. The toolbar at the top includes icons for print, save, email, and other functions. The page content features the title "The Section 108 Study Group Report" in a large, bold, black serif font. Below the title, the text "March 2008" is displayed in a smaller, black serif font. On the left side of the page, there is a vertical sidebar with several icons, including a document icon, a blue document icon, and a green question mark icon. The page is set to page 1 of 212, and the zoom level is 66.7%.

Access via CONTENTdm

The Section 108 Study Group

| | |
|--------------------------|--|
| Title | The Section 108 Study Group report an independent report sponsored by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress. |
| Rights | This work may be protected by copyright. It was originally published on the free Web and has been harvested for preservation purposes under a claim of fair use. Please consult the work itself for additional statements regarding copyright ownership and permissions. Access to this work is provided here for educational purposes only. If you are a copyright owner who objects to the preservation of your work in this fashion, or if you believe that your copyright has been violated by the project's efforts, please make that concern known to The Chesapeake Project, and the work will be removed from the digital archive. |
| Date-Issued | 2008 |
| Subject | Copyright--United States.; Copyright--Electronic information resources--United States.; Fair use (Copyright)--United States.; Copyright and electronic data processing--United States.; Digital media--Law and legislation--United States. |
| Description | Title from title screen (viewed April 9, 2008).; "March 2008.;" Also available in the OCLC Digital Archive. Harvested from http://www.section108.gov/docs/Sec108StudyGroupReport.pdf on May 19, 2008.; Issued with: The Section 108 Study Group report: executive summary |
| Publisher | The Section 108 Study Group |
| Type | Text |
| Format | PDF-1.6; PDF-1.6 |
| URL-Identifier | http://www.section108.gov/docs/Sec108StudyGroupReport.pdf ; http://www.section108.gov/docs/Sec108ExecSum.pdf |
| Language | eng |
| Relation | http://worldcat.org/oclc/221658479/viewonline |
| Format-Extent | 187 p. : digital, PDF file. |
| Relation-Requires | Mode of access: World Wide Web.; System requirements: Adobe Acrobat Reader. |
| OCLC number | 221658479 |

MARC metadata mapped to Qualified Dublin Core

Organizational Framework & Staffing

- Collaborating libraries:
 - State + academic law libraries
 - Different patron groups, mandates
 - Range in staff size from 5, to 15, to 70!
- Requires flexible policies, regular communication

Organizational Framework & Staffing

- Staffing:
 - Directors: project planning, upper-level decision-making, strategy
 - Staff Librarian/Project Coordinators: appointed to coordinate project participation; manage curatorial, cataloging, and digital-archiving tasks
 - Technical Services/Cataloging Librarians: assist with cataloging; consultation on cataloging issues and decisions

Organizational Framework & Staffing

- Time requirements:
 - Georgetown Law Library: 30 hours/week (central project coordinator)
 - Maryland State Law Library: 12 hours/week
 - Virginia State Law Library: 5 hours/week (15 at the beginning of the project)
- Cataloging archived items = most time-consuming aspect of project
 - Majority of archived items represent fugitive or gray literature

Organizational Framework & Staffing

- Flexible Collection Plan developed (template adopted from NDIIPP-sponsored Web-at-Risk project)
 - Describes mission & scope, acquisition/selection methods, metadata policies, access methods, preservation system
- Formal schedule of quarterly meetings to reassess/update project policies and share information as the project progresses through pilot phase

First-Year Project Evaluation

- March 2008: First-year mark of The Chesapeake Project
- Evaluation Parameters:
 - No. of items/titles archived during year one
 - Access statistics
 - Count of archived items altered/removed from original locations on Web
 - Qualitative analysis of challenges & progress

Archiving Activity Figures

- 2,705 digital items harvested & archived, along with preservation metadata, in the OCLC Digital Archive
 - Represents roughly 1,266 titles
- Monthly archiving activity levels varied from 71 to 377 items archived per month
 - Average items archived per month = 225

Access Statistics

- Total access figure = 5,317
 - Access by public (non-authenticated) users = 2,528 (47%)
 - Chesapeake Project access = 2,267 (43%)
 - Other authenticated OCLC libraries and institutions = 533 (10%)
- Monthly access figures ranged
 - Low of 206, July 2007
 - High of 979, September 2007

Archived Items Altered/Missing from Web

- Sample analysis of the original URLs of titles harvested from the Web and archived between the dates of February 27, 2007, and February 29, 2008
- **8.3 percent of the original URLs had become inactive by March 2008!**
 - 10.8 percent of state URLs (state.__.us) inactive
 - 10 percent of government URLs (.gov) inactive
 - 8.3 percent of the organization URLs (.org) inactive

| TOP-LEVEL DOMAIN | TOTAL IN SAMPLE | ACTIVE | INACTIVE | PERCENT INACTIVE |
|------------------|-----------------|--------|----------|------------------|
| .state.__.us | 240 | 214 | 26 | 10.8% |
| .org | 184 | 177 | 7 | 8.3% |
| .gov | 100 | 90 | 10 | 10% |
| .edu | 17 | 15 | 2 | 11.8% |
| .com | 13 | 11 | 2 | 15.4% |
| .net | 11 | 11 | 0 | — |
| .mil | 3 | 3 | 0 | — |
| .us | 3 | 3 | 0 | — |
| .info | 2 | 1 | 1 | 50% |
| .uk | 2 | 2 | 0 | — |
| .au | 1 | 1 | 0 | — |
| .ca | 1 | 1 | 0 | — |
| .int | 1 | 1 | 0 | — |
| [IP address] | 1 | 1 | 0 | — |

The top-level domains of active and inactive original URLs for titles in the sample

Archived Items Altered/Missing from Web (a few samples)

- The future of oil energy security, climate risks, and market opportunities (ceres.org):
<http://worldcat.org/oclc/180862732>
- Briefing report for federal department leadership (ecr.gov):
<http://worldcat.org/oclc/154233375>
- Office of Refugee Resettlement (ORR) annual report to Congress (purl.access.gpo.gov / acf.dhhs.gov):
<http://worldcat.org/oclc/58837769>
- Harford County Juvenile Drug Court performance evaluation final report (courts.state.md.us):
<http://www.worldcat.org/oclc/82366432>

Qualitative Analysis

- First year challenges:
 - Transition to new digital archive system
 - Selection and management of content placed in the digital archive
 - Procedure, policy, and project standard development
 - Learning to use OCLC systems/original cataloging
 - The loss of an influential project leader

Qualitative Analysis

- Accomplishment of project mission/vision:
 - “Participants agree that although great strides have been made in the first year, in order to truly realize the project’s mission and vision, a concerted effort to raise awareness and educate others about the work of The Chesapeake Project will be instrumental in the project’s second and final year”

Project Vision

- The Chesapeake Project aims to set a precedent for a national movement to prevent the widespread loss of legal information in digital formats, securing these materials for generations to come.

Upon reaching the close of its two-year pilot phase in 2009, The Chesapeake Project hopes to help inspire, establish, and galvanize widespread participation in a comprehensive, collaborative, and nationwide preservation program for legal resources.

Next Steps

- Pilot phase ends in 2009
- Aspiration: to evolve into a national Legal Information Archive, shared by participant law libraries throughout the U.S.
- Given the success of the project's first year, and with support of LIPA & AALL, this vision is within reach
- Tackling the challenges of digital preservation represents a means by which law libraries can reclaim their traditional roles as stewards of information in the digital sphere, while also ensuring that our contemporary legal heritage is preserved for generations to come

More Information

- Visit us online:
 - The Chesapeake Project:
- The Legal information Preservation Alliance (LIPA):

legalinfoarchive.org

www.aallnet.org/committee/lipa

Thank You!

- Steve Anderson, Director
Maryland State Law Library
Steve.Anderson@courts.state.md.us
- Gail Warren, State Law Librarian
Virginia State Law Library
gwarren@courts.state.va.us
- Sarah Rhodes, Digital Collections Librarian
Georgetown Law Library
sjr36@law.georgetown.edu