NEW YORK UNIVERSITY

DSpace and Web Material: Inroads and Challenges

Leslie Myrick, NYU

DLF Spring Forum

April 15, 2005

What I’ll Be Covering

- NDIIPP “Web at Risk” Project
- Web Archive Data Object Modeling
- DSpace and HTML
    - Issues, Challenges, Prototypes
- Dspace, METS and Heritrix .arc format
- What’s needed now ; desiderata for DSpace 2.0 +

NDIIPP "Web at Risk" Partnership

- California Digital Library + UC partners
- University of North Texas
- New York University

- SDSC
- Stanford University
- Arizona State University
- Sun Microsystems

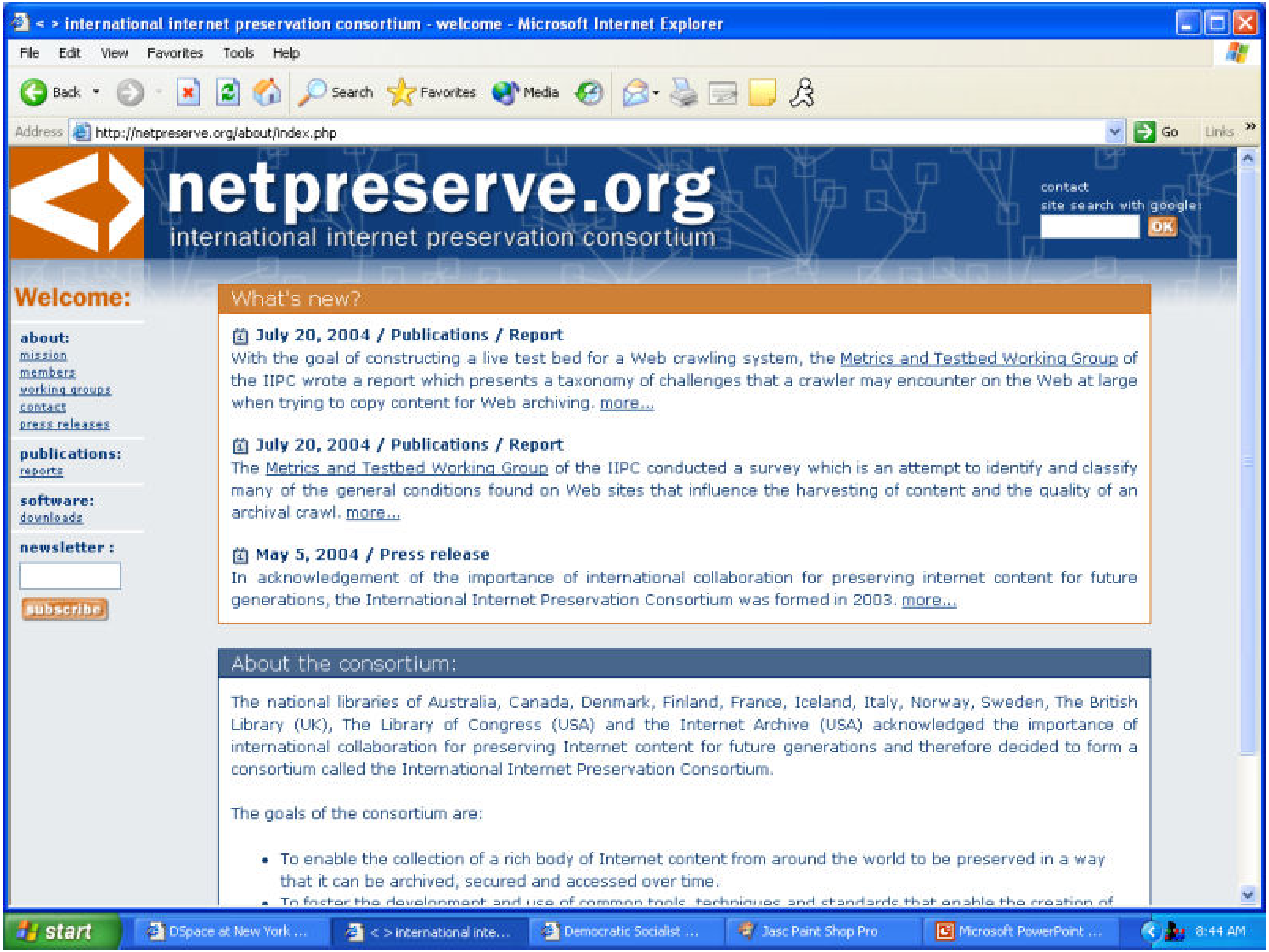
- Our LC AOTR is Martha Anderson

Web Archiving ‘Giants’

- Internet Archive Wayback Machine

- National Library of Australia PANDORA
- Royal Library of Sweden Kulturarw3

- LC MINERVA Project


< > international internet preservation consortium - welcome - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://netpreserve.org/about/index.php
Go
Links
start
DSpace at New York ...
< > international inte...
Democratic Socialist ...
Jasc Paint Shop Pro
Microsoft PowerPoint ...
8:44 AM

netpreserve.org

international internet preservation consortium

contact
site search with google:
OK

Welcome:

about:
mission
members
working groups
contact
press releases

publications:
reports

software:
downloads

newsletter :

subscribe

What's new?

July 20, 2004 / Publications / Report
With the goal of constructing a live test bed for a Web crawling system, the Metrics and Testbed Working Group of the IIPC wrote a report which presents a taxonomy of challenges that a crawler may encounter on the Web at large when trying to copy content for Web archiving. more...

July 20, 2004 / Publications / Report
The Metrics and Testbed Working Group of the IIPC conducted a survey which is an attempt to identify and classify many of the general conditions found on Web sites that influence the harvesting of content and the quality of an archival crawl. more...

May 5, 2004 / Press release
In acknowledgement of the importance of international collaboration for preserving internet content for future generations, the International Internet Preservation Consortium was formed in 2003. more...

About the consortium:

The national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, The British Library (UK), The Library of Congress (USA) and the Internet Archive (USA) acknowledged the importance of international collaboration for preserving Internet content for future generations and therefore decided to form a consortium called the International Internet Preservation Consortium.

The goals of the consortium are:

- To enable the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time.
- To foster the development and use of common tools, techniques and standards that enable the creation of

IIPC

- International Internet Preservation Consortium
  - National Library of Italy (Firenze)
  - Royal Library of Denmark
  - National Library of Finland
  - Internet Archive
  - Royal Library of Sweden
  - National Library of Iceland
  - Library and Archives / Canada
  - National Library of Norway
  - National Library of Australia
  - British Library
  - Library of Congress

Local Expertise/Groundwork

- CDL
  - Web-based Government Information Project
  - California Recall Election Project
- UNT
  - CyberCemetery
- NYU
  - CRL Political Communications Web Archiving Project (Cornell, NYU, Stanford, UT Austin)
- LOC
  - MINERVA
  - IIPC Partnership

Partnership’s Four Paths

- Content Identification, Selection and Acquisition
  - Selection, Curatorial Issues
- Content Harvest and Analysis
  - Crawler
- Content Ingest, Retention and Transfer
  - CDL Digital Preservation Repository + DSpace
  - SRB, other grid technology, other means for synchronized replication
- Partnership Building
  - Technical and Human infrastructures

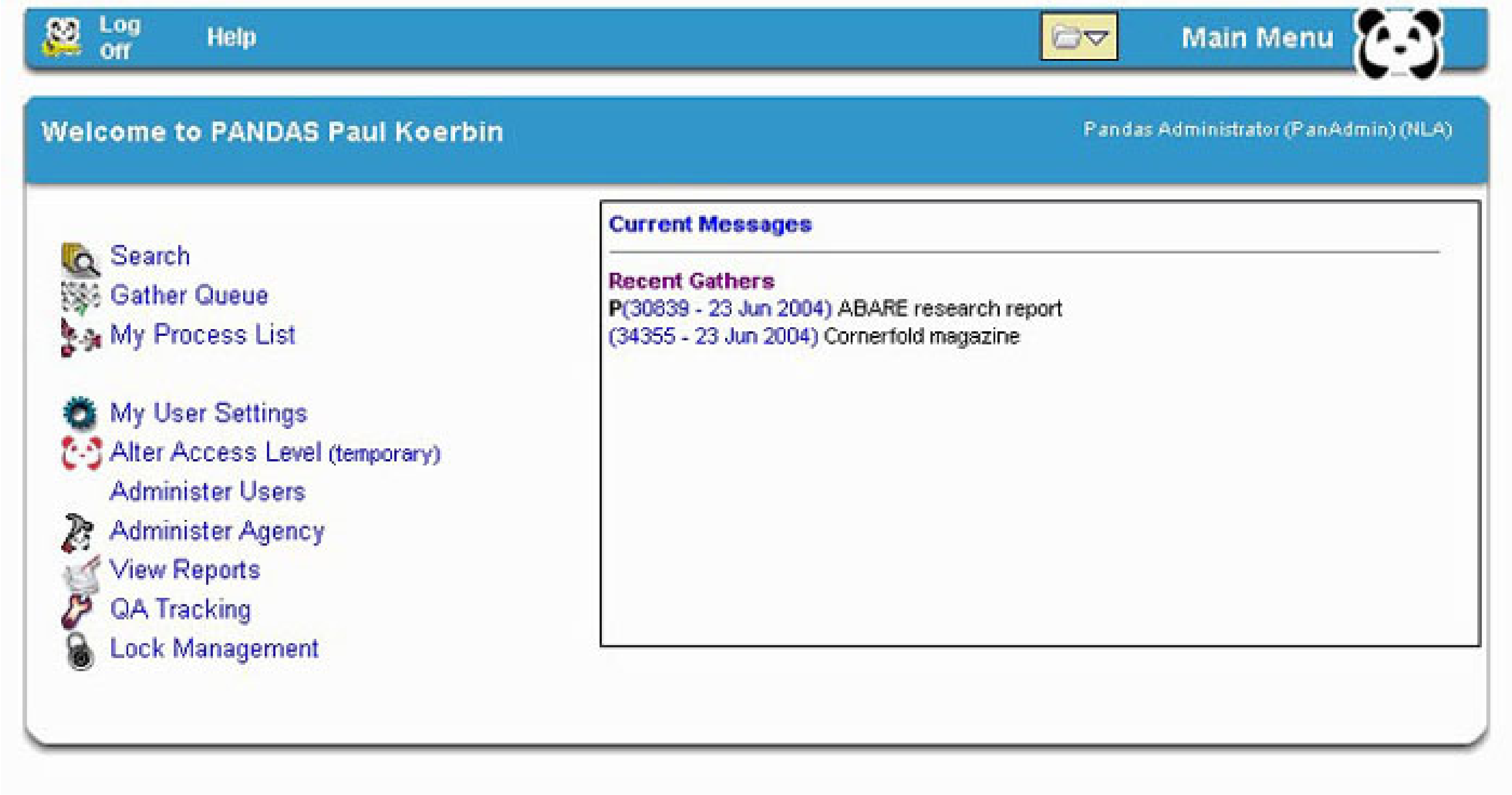
What are we Capturing?

- WADO
  - Web Archive Digital Object
- Crawls, websites?
  - Many seed URLs recursively processed
    - archived websites, however they are defined
- Storage and Archive Format
  - Website mirrors
  - Flat hierarchy: entry page + all files
  - Gzipped archive files (.arc)

Web Capture Tools

- Precrawl / Analysis Tools
  - Some of Linklint's functionality? And more
  - http://www.linklint.org/
    - Analysis of file types; "skipped" actions; missing pages
- Curatorial Interface
  - NLA's PANDAS as one model
  - Andy Boyko's PreCrawl tool for LC / IIPC
  - UIUC/OCLC Echo DEPository tools

PANDAS Interface



Welcome to PANDAS Paul Koerbin

Pandas Administrator (PanAdmin) (NLA)

Current Messages

Recent Gathers

P(30839 - 23 Jun 2004) ABARE research report

(34355 - 23 Jun 2004) Cornerfold magazine

Crawler

- Crawler candidates
  - HTTrack
  - Heritrix

HTTrack

- Developer: Xavier Roche & co.
- http://www.httrack.com/
- Written in C, open source
- Archive format: website mirror; .arc possible
- Core of NLA’s PANDAS application
  - With Java UI modules

Strengths and Weaknesses

- + Configurability
- + Small footprint
- + Incremental crawls possible
- - Scaling issues
- - No multi-machine crawling capability

Heritrix

- Developers: Internet Archive et al
- http://crawler.archive.org/
- Working closely with IIPC
- Java, open-source ; Sourceforge
- Archive format: .arc.gz or mirror
- 1.0 release Aug 2004; up to 1.2.0 (Nov 2004); 1.4.0 (scheduled March 2005)

Heritrix Architecture

- Scope – URIs to process
- Frontier -- controls processing
- Processor chains
  - Prefetch
  - Fetch
  - Extraction
  - Write
  - Postprocess

Strengths and Weaknesses

- + International Standards
- + Programming base / experience
- + Configurability, extensibility
- - Too-sophisticated UI ; reports
- - Memory management
- - No multi-machine processing as yet
- - No incremental harvest capability as yet

Heritrix .arc format (2.x)

- File header for .arc file itself
- Crawl metadata with arc and dc namespacing
  - Crawl host, operator, date
- Each harvested file and its metadata
  - [DNS head]
  - HTTP response headers
  - The captured file

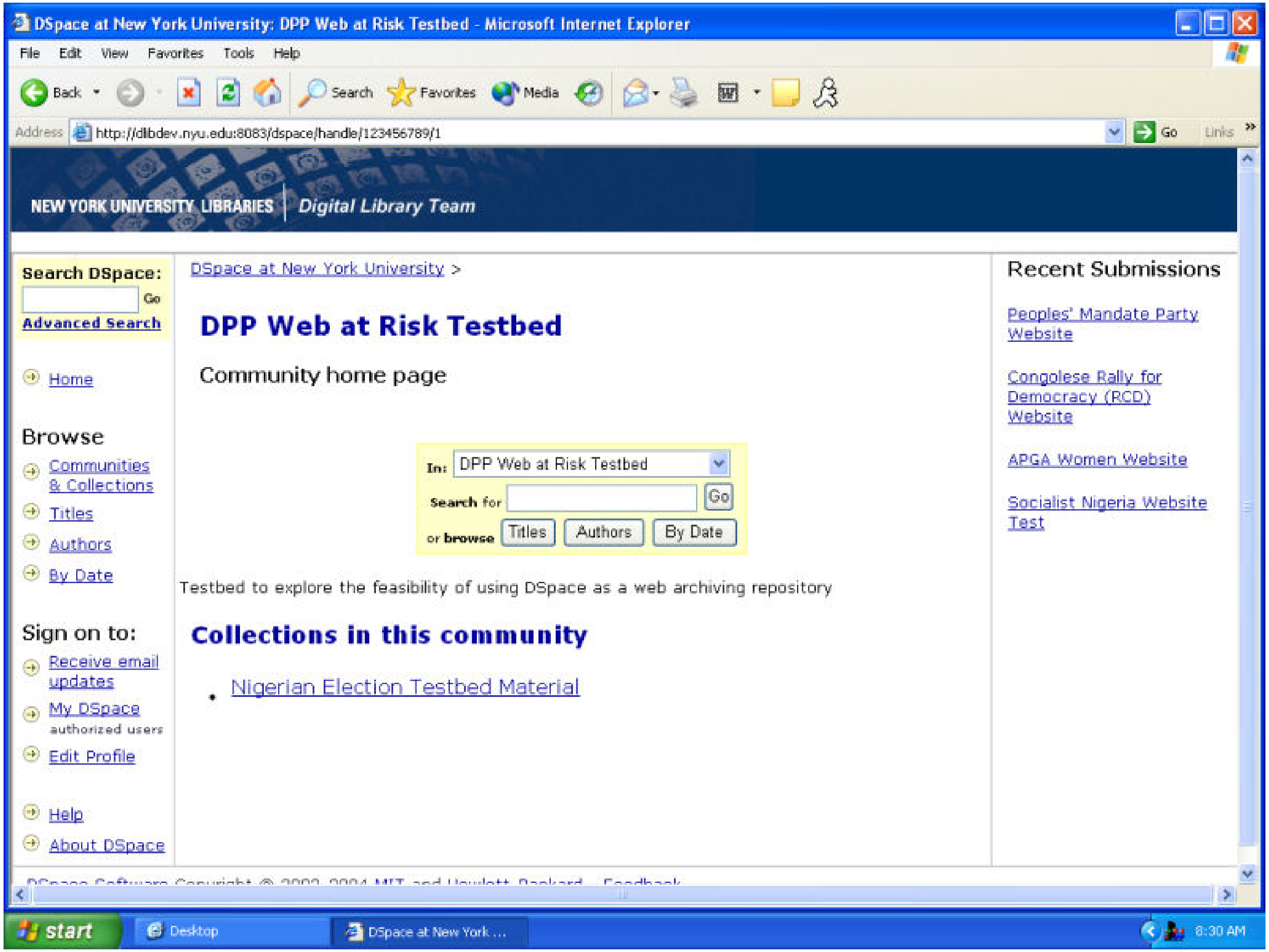
The problem with .arcs

- Many seed URLs per crawl = many sites per .arc
- 100 MB limit (default) = many .arc files per crawl
- Management nightmare?
- Libarc tools promise to mitigate these I/O problems:
  - http://sourceforge.net/projects/libarc/

The other problem with .arcs

- Gzipped format
  - each file and metadata bytestream is a separate member
- Solutions?
  - Interface with Heritrix API
    - can navigate a gzipped .arc:
      - ArcReader methods
      - Create a mini-Wayback machine
  - Use METS to manage transform object vs data object
- Question: How can METS and DSpace be tweaked to handle a zipped web archive?

How does DSpace handle HTML?



DSpace at New York University: DPP Web at Risk Testbed - Microsoft Internet Explorer

Address http://dlibdev.nyu.edu:8083/dspace/handle/123456789/1

NEW YORK UNIVERSITY LIBRARIES | *Digital Library Team*

Search DSpace: Go
Advanced Search

Home

Browse
- Communities & Collections
- Titles
- Authors
- By Date

Sign on to:
- Receive email updates
- My DSpace authorized users
- Edit Profile

- Help
- About DSpace

DSpace at New York University >

DPP Web at Risk Testbed

Community home page

In: DPP Web at Risk Testbed
Search for Go
or browse Titles Authors By Date

Testbed to explore the feasibility of using DSpace as a web archiving repository

Collections in this community

- Nigerian Election Testbed Material

Recent Submissions

Peoples' Mandate Party Website

Congolese Rally for Democracy (RCD) Website

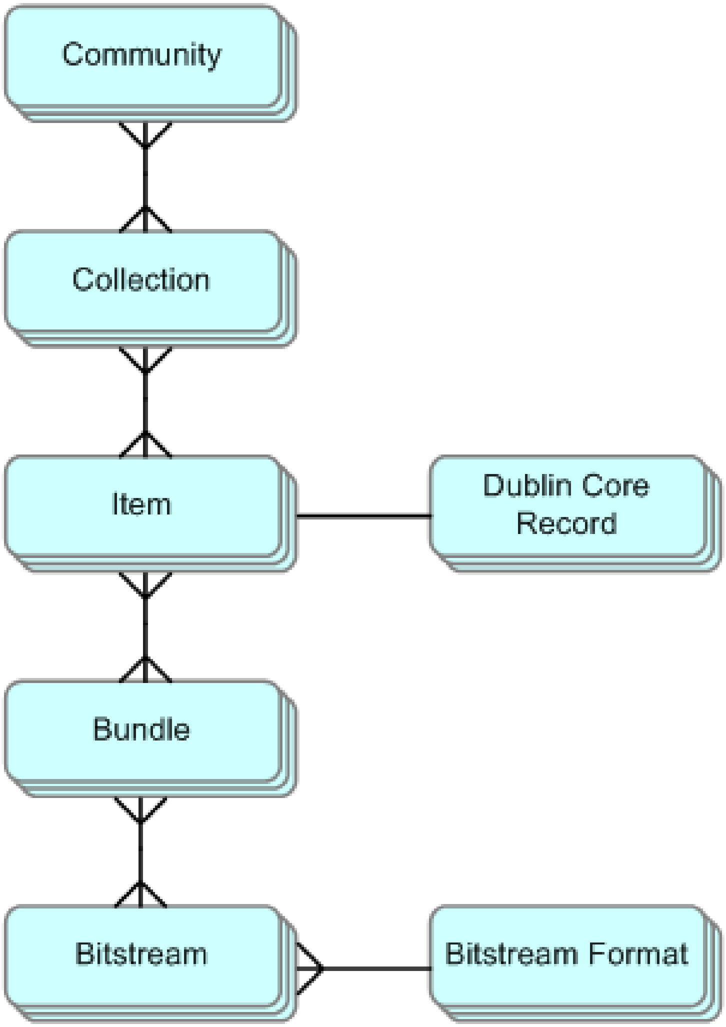
APGA Women Website

Socialist Nigeria Website Test

DSpace Software Copyright © 2002-2004 MIT and Hewlett Packard - Feedback

File Edit View Favorites Tools Help
Back Search Favorites Media
Go Links
start
Desktop
DSpace at New York ...
8:30 AM

DSpace Data Model



Objects in DSpace Data Model

- Item

  A technical report; a data set with accompanying description; a video recording of a lecture

- Bundle

  A group of HTML and image bitstreams making up an HTML document

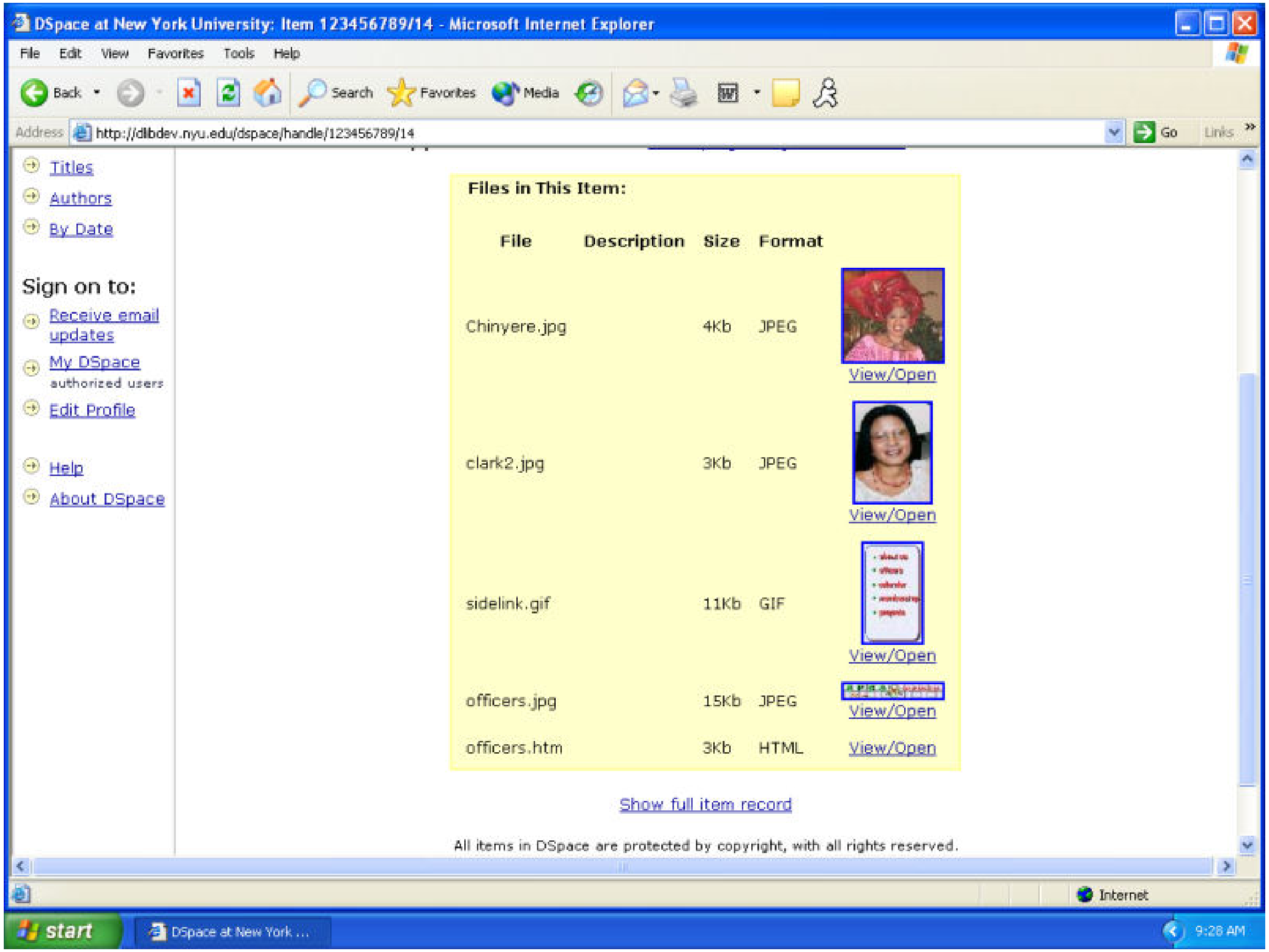
- Bitstream

  A single HTML file; a single image file; a source code file

DSpace and HTML: Webpage as Item

- Webpage as Item with its own handle
  - a bundle of files with a nominated primary bitstream (HTML)
- Integrated view of HTML page
  - together with all other files necessary for rendering (css, images, javascript).

HTML Bundle before …


DSpace at New York University: Item 123456789/14 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
Go
Links
Internet
start
DSpace at New York ...
9:28 AM

http://dlibdev.nyu.edu/dspace/handle/123456789/14

- Titles
- Authors
- By Date

Sign on to:

- Receive email updates
- My DSpace
authorized users
- Edit Profile
- Help
- About DSpace

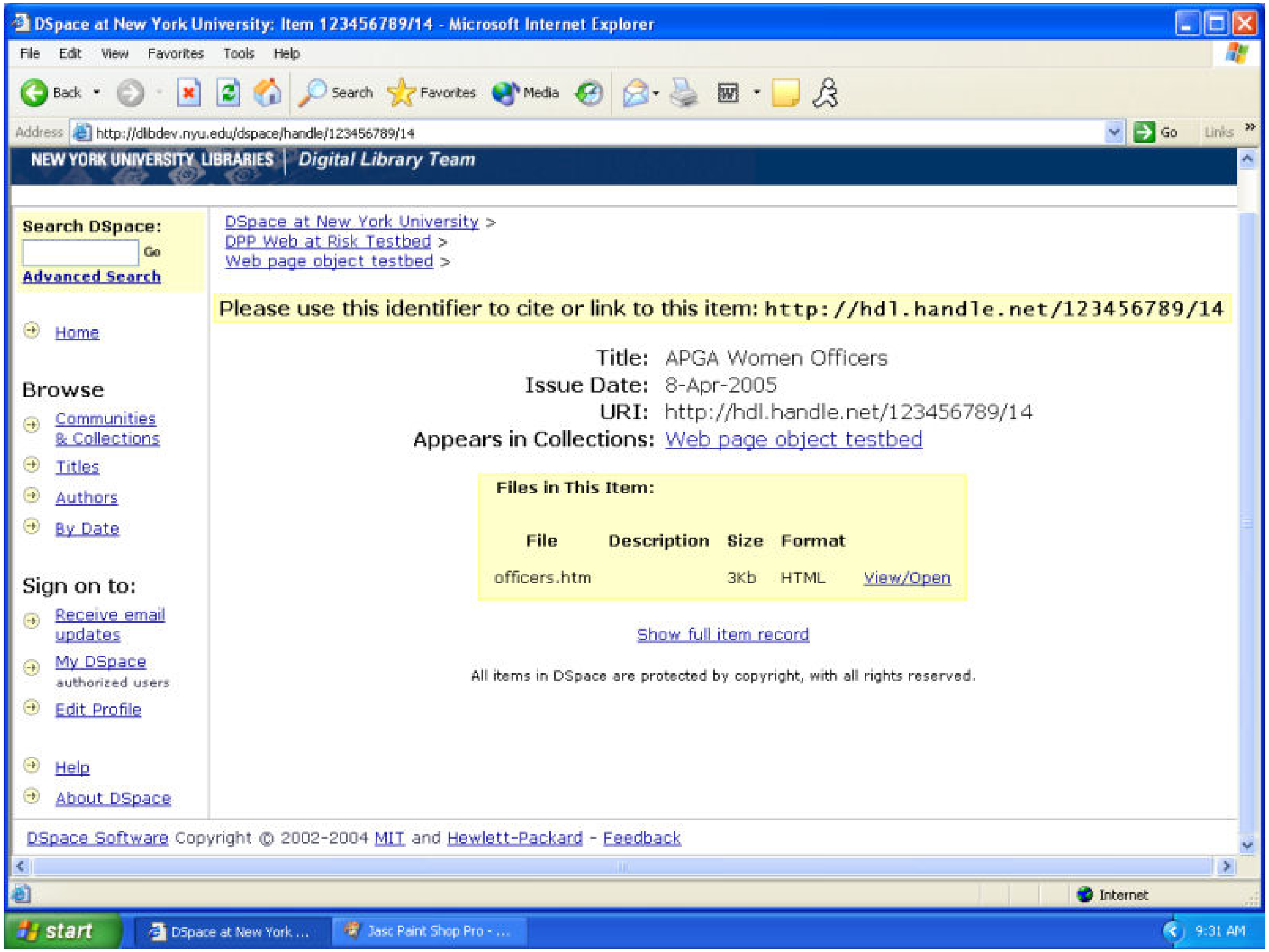
Files in This Item:

File	Description	Size	Format	
Chinyere.jpg		4Kb	JPEG	View/Open
clark2.jpg		3Kb	JPEG	View/Open
sidelink.gif		11Kb	GIF	View/Open
officers.jpg		15Kb	JPEG	View/Open
officers.htm		3Kb	HTML	View/Open

Show full item record

All items in DSpace are protected by copyright, with all rights reserved.

… HTML Bundle after


DSpace at New York University: Item 123456789/14 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://dlibdev.nyu.edu/dspace/handle/123456789/14
Go
Links
NEW YORK UNIVERSITY LIBRARIES | Digital Library Team
Internet
start
DSpace at New York ...
Jasc Paint Shop Pro - ...
9:31 AM

Search DSpace:
Go
Advanced Search

Home

Browse
- Communities & Collections
- Titles
- Authors
- By Date

Sign on to:
- Receive email updates
- My DSpace authorized users
- Edit Profile

Help
About DSpace

DSpace at New York University >
DPP Web at Risk Testbed >
Web page object testbed >

Please use this identifier to cite or link to this item: `http://hdl.handle.net/123456789/14`

Title:	APGA Women Officers
Issue Date:	8-Apr-2005
URI:	http://hdl.handle.net/123456789/14
Appears in Collections:	Web page object testbed

Files in This Item:

File	Description	Size	Format	
officers.htm		3Kb	HTML	View/Open

Show full item record

All items in DSpace are protected by copyright, with all rights reserved.

DSpace Software Copyright © 2002-2004 MIT and Hewlett-Packard - Feedback

Webpage-Level Item + and -

- + Search returns hits at page-level
- - Can’t navigate natively within DSpace to any other webpage items from the same site that might be in the repository.
  - Could use METS DIP to navigate handles
- - Requires iterative loads of same files from other pages in same website (css, icons used across a site)

Website as Item

- Website as Item
  - bundle of files with nominated primary bitstream (entry page's HTML).
- All other files (HTML and non-HTML) flattened out beneath this primary bitstream.

Website-Level Item + and -

- + Files ingested and stored once
- + Can navigate hyperlink structure natively to DSpace
  - with some adjustments to archived HTML files.
- - Search returns hits at website level
- - HTML files’ hyperlinks require some tweaking before loading

DSpace Restrictions on HTML Objects

- No dynamic content (e.g. PHP, CGI)
- * All files must have unique names
- All hyperlinks must be relative and not refer to parents
  - myfile.html is okay
  - /myfile.html is not okay
  - Originally document-relative paths were anathema (../myfile.html)

All Filenames Must be Unique

- Problematic in “real world” archiving
  - – index.html at every directory level in site
- Use long filenames = path + filename

```
<snip>
paper/2003/feb-mar/SDFebMar03.rtf
paper/index.htm
statements/index.html
statements/nlc.html
students/index.html
students/war.html
index.html
</snip>
```

Sidebar: Websites in DSpace:CWSpace

- William Reilly and Rob Wolfe (kudos and thanks)
- http://cwspace.mit.edu/
- Archiving MIT Open Courseware in DSpace
- IMS-CP packaging; conversion to METS AIP
- Impetus for changes to HTMLServlet and elsewhere; recent patch (11435750)
  - Larry Stone

Recent Servlet Revisions

- Slashes in filenames inherently problematic
- Long filenames in contents file for ItemImporter can now be:
  - Handled by bitstreamServlet
  - Rendered by HTMLServlet as distinct items:

```
<snip>
paper/index.htm
statements/index.html
students/index.html
index.html
</snip>
```

Adjustments to HTML files

- Document-relative paths within the files' hyperlinks now mandatory below root level (../)
- Necessary to navigate relative to what HTTPServlet perceives as "server root":
  - "server root" = website item handle
  - Concatenate long filename to it:
  - http://dspace.myu.edu/12345678/9/students/index.html

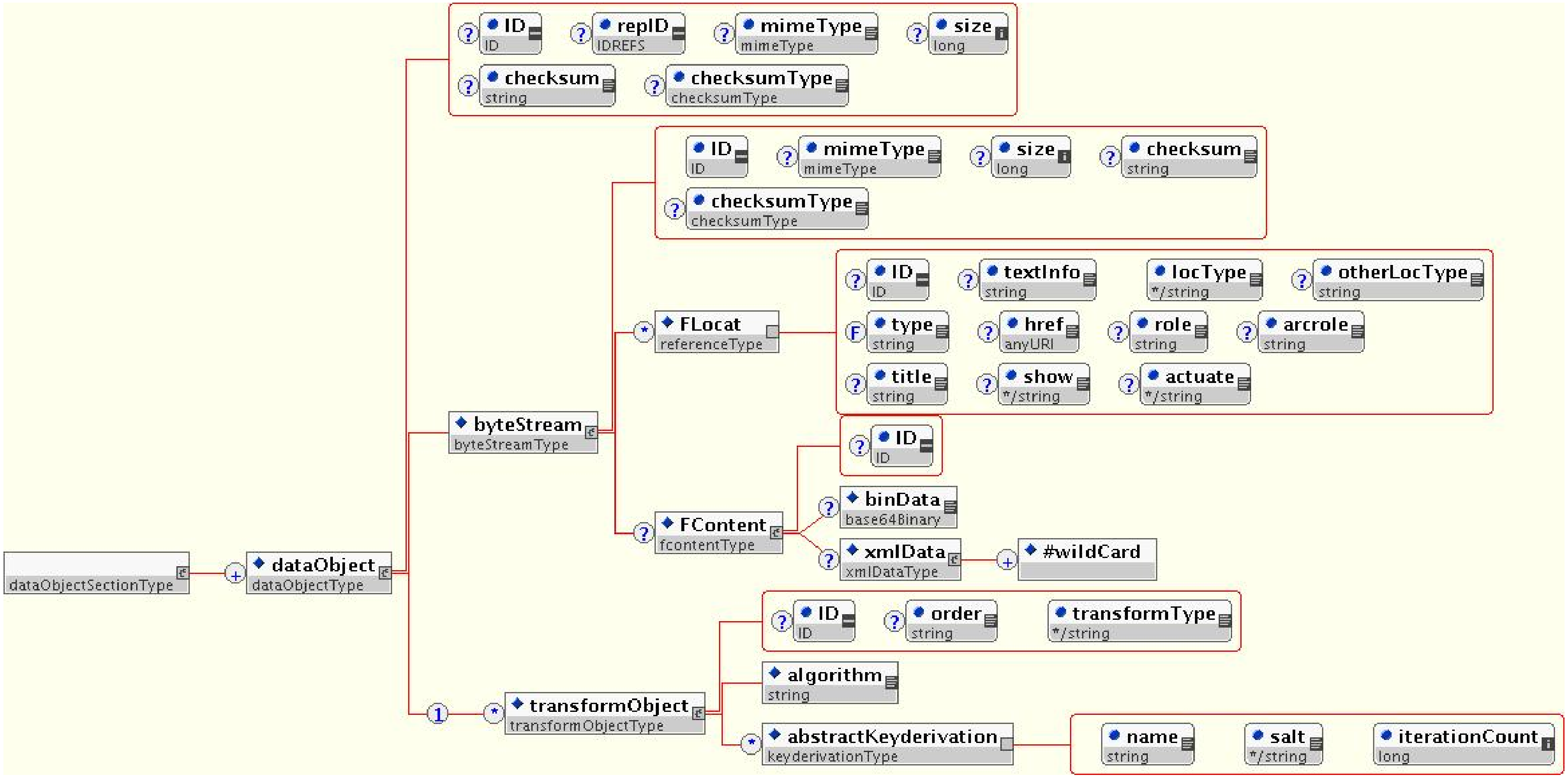
DSpace and Heritrix .arc.gz: Problem and Solutions?

- Problem: .arc.gz can be stored but not accessed as such; invokes gunzip, stuffit etc.
- Solutions?
  - Unzipped .arc could be manipulated locally by third party XSLT, e.g. METS Viewer
  - Interface with Heritrix API or with other tools that can handle .arc.gz format, e.g. ArcReader methods
- Caveat
  - Have to manage transform object over against data object

DataObject vs transformObject

- XFDU dataObject vs transformObject
  - Transform nested in the dataObject

XFDU DataObject and TransformObject



Basic METS Website Data Model

```
<div> website
    <div> HTML page
        <fptr>
            <par>
                <area> bitstream (HTML)
                <area> bitstream (IMG)
            </par>
        </fptr>
        <div/> hyperlink on HTML page
    </div>
    …
</div>
```

METS structMap for webpage

```
<METS:div DMDID="DM1" TYPE="web page" ID="page18"
   LABEL="www.apgawomen.org/index.html">
      <METS:fptr>
          <METS:par>
               <METS:area FILEID="FID18"/>                [index.html ]
               <METS:area FILEID="FID1036"/>              [notjust.swf]
               <METS:area FILEID="FID1043"/>              [apgawnew.swf]
               <METS:area FILEID="FID1075"/>              [enterarrow.gif]
              </METS:par>
        </METS:fptr>
      <METS:div TYPE="hyperlink" ID="LINK1" LABEL="home">
              <METS:fptr>
                  <METS:area BEGIN="000" BETYPE="BYTE" END="111"
                     FILEID="FID18"/>
              </METS:fptr>
      </METS:div>
```

Mapping Hyperlink Structure

```
<METS:div DMDID="DM1" TYPE="web page" ID="page18" LABEL=“www.apgawomen.org/index.html">
       <METS:fptr>
              <METS:par>
                 <METS:area FILEID="FID18"/>                [index.html ]
                 <METS:area FILEID="FID1036"/>              [notjust.swf]
                 <METS:area FILEID="FID1043"/>              [apgawnew.swf]
                 <METS:area FILEID="FID1075"/>              [enterarrow.gif]
              </METS:par>
        </METS:fptr>
        <METS:div TYPE="hyperlink" ID="LINK1" LABEL="home">
            <METS:fptr>
               <METS:area BEGIN="000" BETYPE="BYTE" END="111" FILEID="FID18"/>
            </METS:fptr>
         </METS:div>
</METS:div>
. . .
<METS:structLink>
   <METS:smLink from="LINK1" to="page1059" xlink:title="home"/>
   <METS:smLink from="LINK2" to="page113" xlink:title=”officers"/>
   <METS:smLink from="LINK3" to="page102" xlink:title=”calendar"/>
</METS:structLink>
```

Possible adjustments to METS for .arc.gz

- Nested <file>s in <fileSec>
- <stream>s to handle metadata headers

- Possibly used along with either:
  - Nested transformObject for each file or
  - Dual structMaps
    - One for transform object
      - SM01: gzip object as the root div; .arc as child; all files as grandchildren;
    - One for website object unzipped
      - SM02: standard logical representation of files as they would exist outside of .arc.gz

One possibility: Nested <file>s

```
<METS:fileSec>
   <METS:fileGrp>
      <METS:file ID="FID1" MIMETYPE="application/x-gzip" ADMID="ADM001">
         <METS:FLocat LOCTYPE="OTHER" OTHERLOCTYPE= "TRANSFORM-URI"
xlink:href="file:///usr/local/heritrix/jobs/Test07/arcs/IAH-20050203191213-00000-
euterpe.bobst.nyu.edu.arc.gz"></METS:FLocat>
               <METS:file ID="FID2" MIMETYPE="text/plain" ADMID="ADM01">
                  <METS:FLocat LOCTYPE="OTHER" OTHERTYPE="TRANSFORM-URI"
xlink:href="file://usr/local/heritrix/jobs/Test07/arcs/IAH-20050203191213-00000-
euterpe.bobst.nyu.edu.arc"></METS:FLocat>
               <METS:file ID="FID3" MIMETYPE=" text/html" ADMID="ADM1">
                  <METS:FLocat LOCTYPE="URL" xlink:href="www.apgawomen.org/"></METS:FLocat>
               </METS:file>
               [ other website members / files here ]
             </METS:file>
       </METS:file>
   </METS:fileGrp>
</METS:fileSec>
```

METS StructMap01: Transform Object

```
<METS:structMap ID="SM01" TYPE="logical-transformation">
   <METS:div DMDID="DM01" TYPE="web site" ID="page1" LABEL="www.apgawomen.org/">
      <METS:fptr>
             <METS:par>
               <METS:area BEGIN="1725" BETYPE="BYTE" FILEID="FID1"></METS:area>
               <METS:area BEGIN="6571" BETYPE="BYTE" FILEID="FID1"></METS:area>
               <METS:area BEGIN="2670" BETYPE="BYTE" FILEID="FID1"></METS:area>
               <METS:area BEGIN="17561" BETYPE="BYTE" FILEID="FID1"></METS:area>
            </METS:par>
      </METS:fptr>
      <METS:div TYPE="hyperlink" ID="LINK1" LABEL="home">
         <METS:fptr>
                <METS:area BEGIN="000" BETYPE="BYTE" FILEID="FID3"></METS:area>
             </METS:fptr>
      </METS:div>
          . . .
   </METS:div>
</METS:structMap>
```

METS StructMap02 Unzipped Website Object

```
     <METS:structMap ID="SM02" TYPE="logical">
        <METS:div DMDID="DM01" TYPE="web site" ID="page18" LABEL="www.apgawomen.org/">
           <METS:fptr>
                  <METS:par>
                     <METS:area FILEID="FID3"></METS:area>
                         <METS:area FILEID="FID1036"></METS:area>
                         <METS:area FILEID="FID1043"></METS:area>
                         <METS:area FILEID="FID1075"></METS:area>
                  </METS:par>
               </METS:fptr>
               <METS:div TYPE="hyperlink" ID="LINK1" LABEL="home">
                  <METS:fptr>
      <METS:area BEGIN="000" BETYPE="BYTE" END="111" FILEID="FID3"></METS:area>
                  </METS:fptr>
                </METS:div>
        . . .
   </METS:div>
</METS:structMap>
```

DSpace Desiderata: Ingest

- Load Scripts (Website SIP Client):
  - Script to visit files in load directory and write contents page for ItemImport
  - Script to automate dublin_core.xml
  - Script to "correct" hyperlinks in HTML
    - All pages siblings below root with ../ path
    - Disable mailtos, cgis, etc.
    - Archive or correct links to external pages
- Functionality to automate nomination of primary bitstream
- METS Import
  - Finer-grained metadata possible

DSpace Desiderata: Archival Storage

- Functionality to manage version control issues
  - Successive snapshots of same site
  - Migration of bitstreams
- Facilitate cross-collection item linking

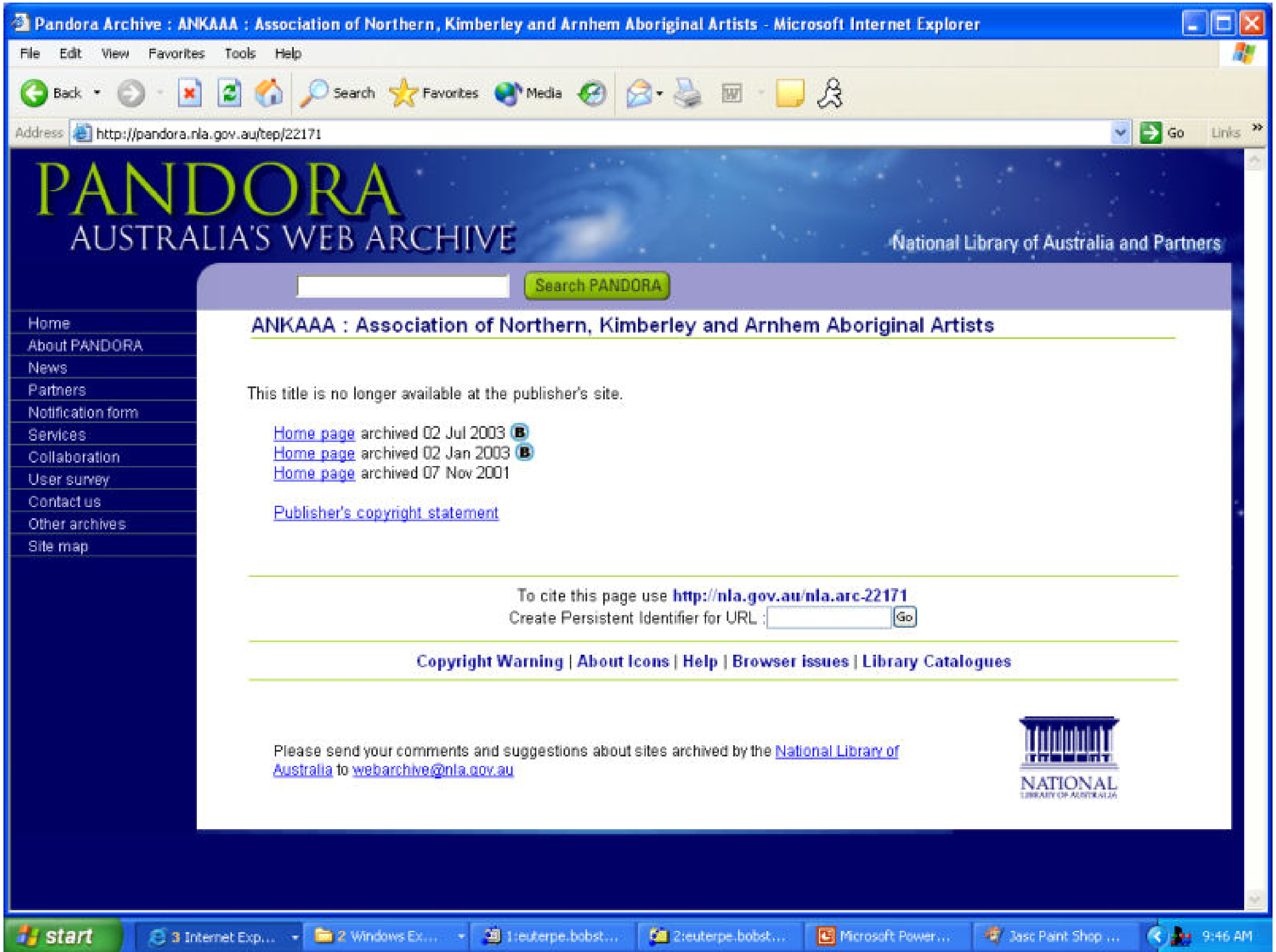
DSpace Desiderata: Metadata

- Ability to apply (and discover) more metadata at bitstream level
  - Controlled descriptive information for HTML, PDF, etc
  - Extend dublin_core.xml SIP to bitstream level?
  - Develop METS SIP?
- METS AIP
  - First of all, a structMap (or two)
  - Manage transform objects' (.arc, .gz) relationship to data objects (zipped files)
  - Reflect complexity of hyperlinked object
    - structLink cross-reference to structMap

DSpace Desiderata: Access

- Display search results at bitstream level
  - for HTML, PDF, MSWord, etc
- Manage navigation of different versions of same site
  - Title Entry Page (NLA PANDAS)
  - IA Resource Page (Wayback, MINERVA)
  - Timeline (NWA Toolset)

PANDAS TEP Interface



ANKAAA : Association of Northern, Kimberley and Arnhem Aboriginal Artists

This title is no longer available at the publisher's site.

Home page archived 02 Jul 2003 B
Home page archived 02 Jan 2003 B
Home page archived 07 Nov 2001

Publisher's copyright statement

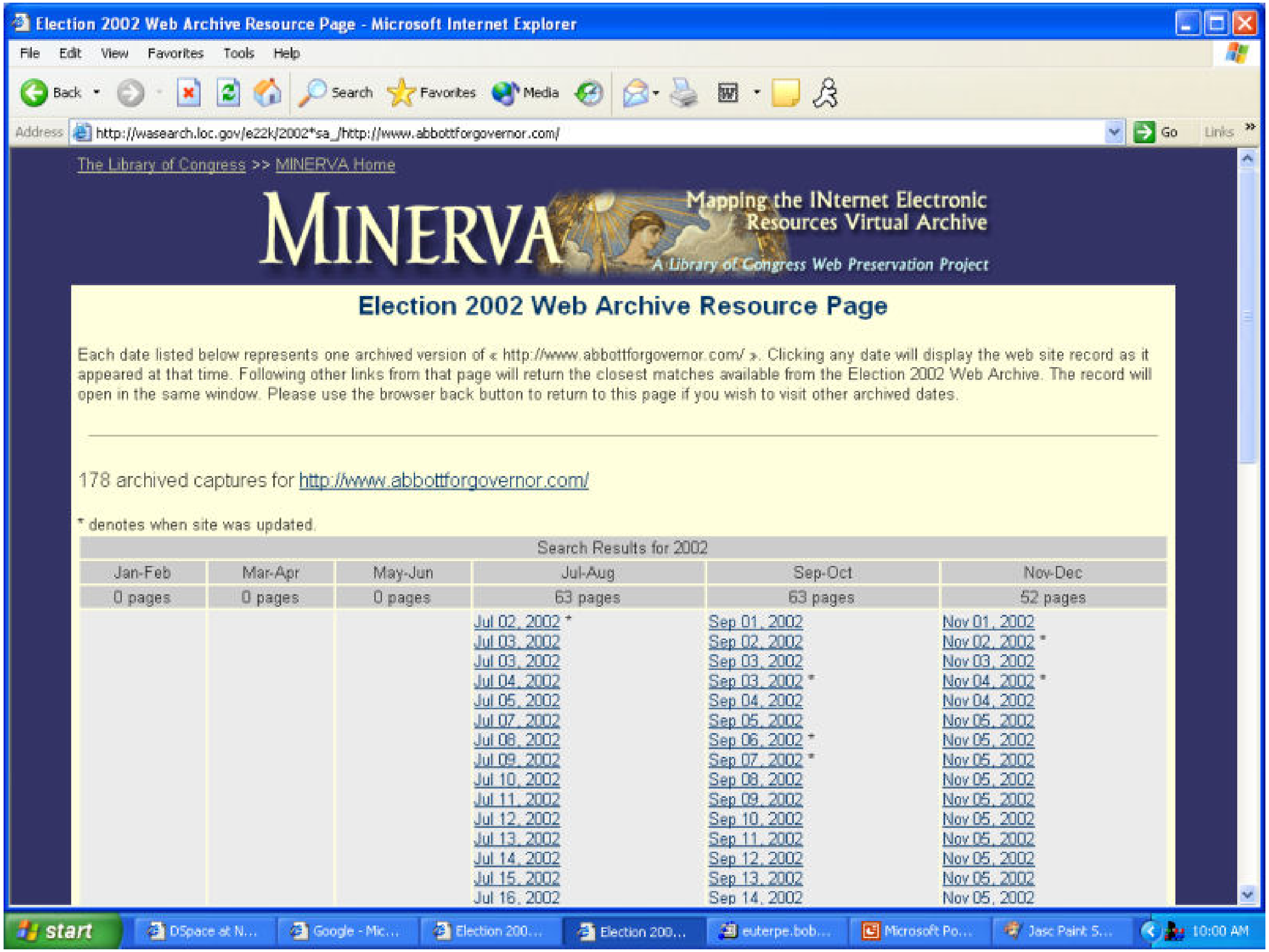
To cite this page use **http://nla.gov.au/nla.arc-22171**
Create Persistent Identifier for URL : Go

Copyright Warning | About Icons | Help | Browser issues | Library Catalogues

Please send your comments and suggestions about sites archived by the National Library of Australia to webarchive@nla.gov.au

NATIONAL LIBRARY OF AUSTRALIA

IA /MINERVA Resource Page



http://wasearch.loc.gov/e22k/2002*sa_/http://www.abbottforgovernor.com/

The Library of Congress >> MINERVA Home

MINERVA Mapping the INternet Electronic Resources Virtual Archive

A Library of Congress Web Preservation Project

Election 2002 Web Archive Resource Page

Each date listed below represents one archived version of « http://www.abbottforgovernor.com/ ». Clicking any date will display the web site record as it appeared at that time. Following other links from that page will return the closest matches available from the Election 2002 Web Archive. The record will open in the same window. Please use the browser back button to return to this page if you wish to visit other archived dates.

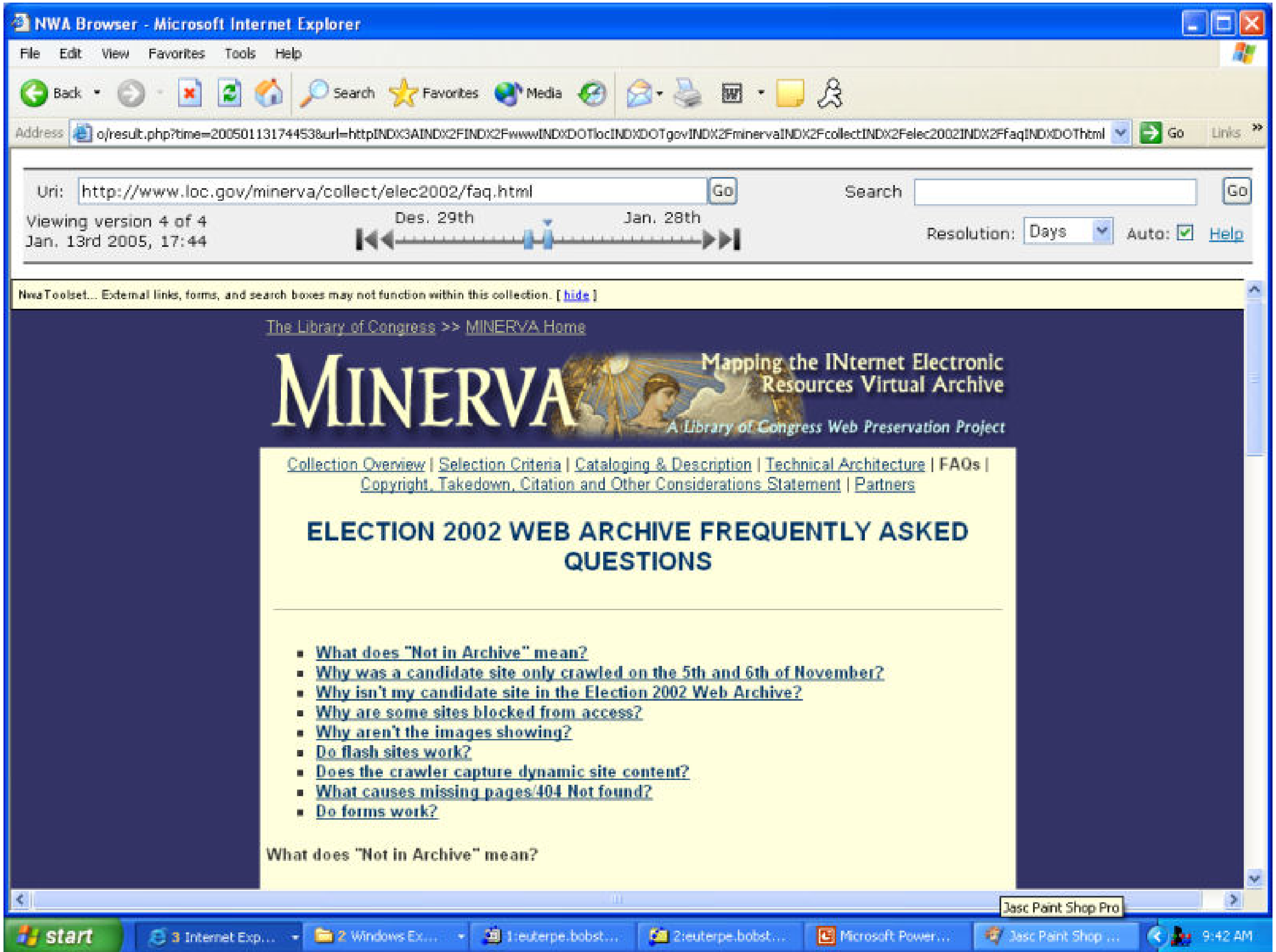
178 archived captures for http://www.abbottforgovernor.com/

* denotes when site was updated.

Search Results for 2002

Jan-Feb	Mar-Apr	May-Jun	Jul-Aug	Sep-Oct	Nov-Dec
0 pages	0 pages	0 pages	63 pages	63 pages	52 pages
			Jul 02, 2002 *	Sep 01, 2002	Nov 01, 2002
			Jul 03, 2002	Sep 02, 2002	Nov 02, 2002 *
			Jul 03, 2002	Sep 03, 2002	Nov 03, 2002
			Jul 04, 2002	Sep 03, 2002 *	Nov 04, 2002 *
			Jul 05, 2002	Sep 04, 2002	Nov 04, 2002
			Jul 07, 2002	Sep 05, 2002	Nov 05, 2002
			Jul 08, 2002	Sep 06, 2002 *	Nov 05, 2002
			Jul 09, 2002	Sep 07, 2002 *	Nov 05, 2002
			Jul 10, 2002	Sep 08, 2002	Nov 05, 2002
			Jul 11, 2002	Sep 09, 2002	Nov 05, 2002
			Jul 12, 2002	Sep 10, 2002	Nov 05, 2002
			Jul 13, 2002	Sep 11, 2002	Nov 05, 2002
			Jul 14, 2002	Sep 12, 2002	Nov 05, 2002
			Jul 15, 2002	Sep 13, 2002	Nov 05, 2002
			Jul 16, 2002	Sep 14, 2002	Nov 05, 2002

NWA Toolset Timeline Access


NWA Browser - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address
Go Links

o/result.php?time=20050113174453&url=httpINDX3AINDX2FINDX2FwwwINDXDOTlocINDXDOTgovINDX2FminervaINDX2FcollectINDX2Felec2002INDX2FfaqINDXDOThtml

Uri: http://www.loc.gov/minerva/collect/elec2002/faq.html Go

Search Go

Viewing version 4 of 4
Jan. 13rd 2005, 17:44

Des. 29th Jan. 28th

Resolution: Days Auto: ☑ Help

NwaToolset... External links, forms, and search boxes may not function within this collection. [hide]

The Library of Congress >> MINERVA Home

MINERVA Mapping the INternet Electronic Resources Virtual Archive

A Library of Congress Web Preservation Project

Collection Overview | Selection Criteria | Cataloging & Description | Technical Architecture | FAQs | Copyright, Takedown, Citation and Other Considerations Statement | Partners

ELECTION 2002 WEB ARCHIVE FREQUENTLY ASKED QUESTIONS

- What does "Not in Archive" mean?
- Why was a candidate site only crawled on the 5th and 6th of November?
- Why isn't my candidate site in the Election 2002 Web Archive?
- Why are some sites blocked from access?
- Why aren't the images showing?
- Do flash sites work?
- Does the crawler capture dynamic site content?
- What causes missing pages/404 Not found?
- Do forms work?

What does "Not in Archive" mean?

DSpace < > Heritrix

- Use METS for supplemental management of transform vs unzipped object
- Interface with Heritrix API
    - To write SIPs from the .arc
    - For possible creation of mirrors
        - MirrorWriterProcessor
    - For Wayback functionality
        - ArcReader

For More Information

- leslie@nyu.edu
- http://dlibdev.nyu.edu/demo/DLF2005.html
- DPPWAR/ NYU DSpace Testbed
  - http://dlibdev.nyu.edu/dspace/handle/123456789/1

Go to Demo