Archive-It and CINCH tool:

Using web harvesting to facilitate born-digital preservation

Kathleen Kenney

Archive-It Partners Meeting 2012

State Library of North Carolina

Collection: State government, genealogy, North Caroliniana

Support: All North Carolina libraries

Services & Materials: for those with visual/physical disabilities

State Library of North Carolina

Collection: State government, genealogy, North Caroliniana

Support: All North Carolina libraries

Services & Materials: for those with visual/physical disabilities

North Carolina General Statute 125

The State Library shall be the official, complete, and permanent depository for all State publications…

- Session laws
- Annual reports
- Technical reports
- Newsletters
- Websites
- and more …

North Carolina General Statute 125

The State Library shall be the official, complete, and permanent depository for all State publications…

Challenges

Manual collection

- We're not getting it all
- The ingested object may not be "authentic"
- We have to ~~badger~~ encourage contributors

Website archiving

- A web archive is hard for users to understand
- We can't provide the continuity from digitized to digital
- We have tons o' data

North Carolina State Government Web Site Archive

Semiannual (ID #20120918121210545)

Crawl Report

Started: September 18, 2012 8:12:11 AM
Completed: September 25, 2012 1:23:12 PM

<< Back to Reports

Scope-It Crawl Explorer

Summary Hosts Seed Status Seed Source File Types PDFs Videos QA

PDFs

Download

The PDF report lists all the PDF files that were crawled (including both new and unchanged PDFs). This information is available 4 hours after a crawl completes. If it has already been more than 4 hours since the crawl completed, then there were no PDFs discovered in this crawl.

« First « Previous Next » Last » 1 through 100 of 250281

URL		
http://67.199.66.38/uploads/File/education/curriculum/A%20house%20for%20hermit%20crab.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Cool%20Crabs.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/ECO-Logical.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Exploring%20the%20Estuary.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Fish%20forms.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Ocean%20movements.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Salt%20Marsh%20Players.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Shore%20Profiles.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Water,%20Water%20Everywhere.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Directions%20for%20Flip%20Foldable_0.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Directions%20for%20Windowpane%20Foldable.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Rainforest%20Travel%20Brochure%20and%20Rubric.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Rainforest%20Virtual%20Field%20Trip_0.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Types%20of%20Clouds.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/Breast_Self_Exam(english-spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/active.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/breast_factFINAL(spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/breast_factFINALjan.08.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/carbohydrates.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/cervical_factFINAL(spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/cervical_factFINALjan.08.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/fats.pdf	View All Captures	Metadata

Our Use Case

How can we extract, use, & preserve the publications discovered by our web archiving project, in an automated and preservation-responsible way?

Preservation Storage

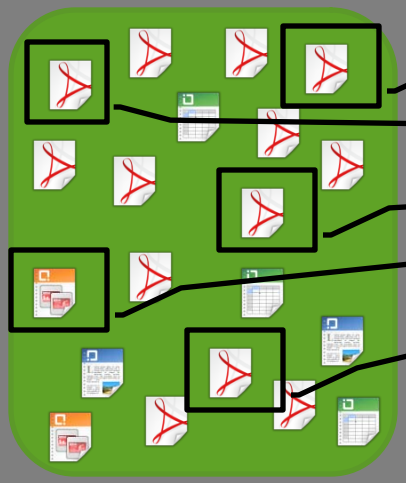
CINCH (Capture INgest CHecksum) is a tool that automates the transfer of online content to a repository, using ingest technologies appropriate for digital preservation.

More familiarly, CINCH

- grabs freely available online content,
- authenticates it,
- extracts metadata, and
- readies it for repository ingest.

- Modular
- Flexible
- Easy to use
- Repository-neutral
- Open source
- Hosted (North Carolina institutions)

The Web



File Types

- gif, jpg, png
- Microsoft Excel Powerpoint Word
- mp3
- pdf
- txt, csv

CINCH

- Checksums
- De-duplication
- Virus scan
- Metadata extraction
- Audit trail

What you get

- Descriptive & administrative metadata
- Audit trail
- Error list

North Carolina State Government Web Site Archive

Semiannual (ID #20120918121210545)

Crawl Report

Started: September 18, 2012 8:12:11 AM
Completed: September 25, 2012 1:23:12 PM

<< Back to Reports

Scope-It Crawl Explorer

Summary Hosts Seed Status Seed Source File Types PDFs Videos QA

PDFs

Download

The PDF report lists all the PDF files that were crawled (including both new and unchanged PDFs). This information is available 4 hours after a crawl completes. If it has already been more than 4 hours since the crawl completed, then there were no PDFs discovered in this crawl.

« First « Previous Next » Last » 1 through 100 of 250281

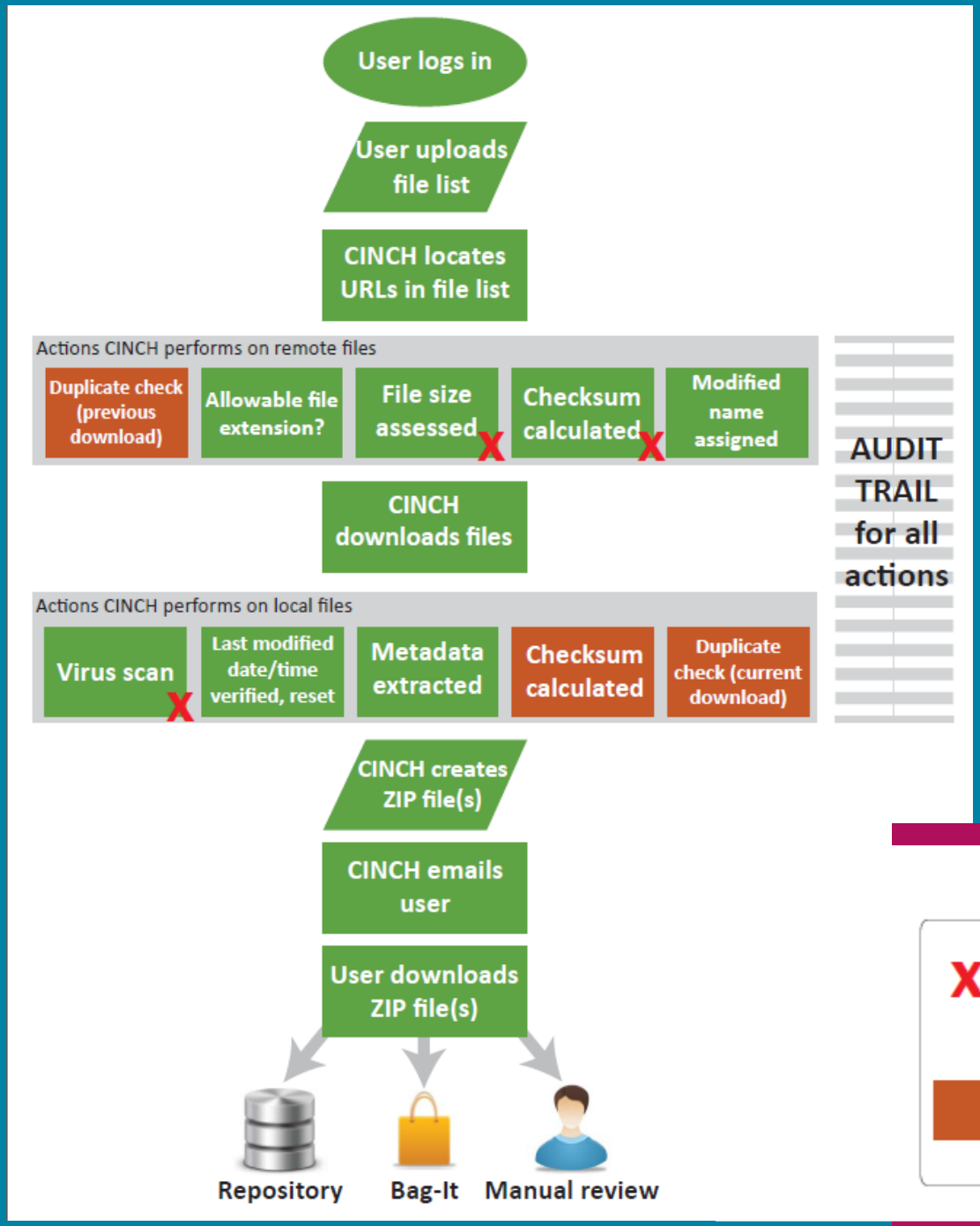
URL		
http://67.199.66.38/uploads/File/education/curriculum/A%20house%20for%20hermit%20crab.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Cool%20Crabs.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/ECO-Logical.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Exploring%20the%20Estuary.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Fish%20forms.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Ocean%20movements.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Salt%20Marsh%20Players.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Shore%20Profiles.pdf	View All Captures	Metadata
http://67.199.66.38/uploads/File/education/curriculum/Water,%20Water%20Everywhere.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Directions%20for%20Flip%20Foldable_0.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Directions%20for%20Windowpane%20Foldable.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Rainforest%20Travel%20Brochure%20and%20Rubric.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Rainforest%20Virtual%20Field%20Trip_0.pdf	View All Captures	Metadata
http://artnc.org/sites/default/files/documents/Types%20of%20Clouds.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/Breast_Self_Exam(english-spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/active.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/breast_factFINAL(spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/breast_factFINALjan.08.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/carbohydrates.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/cervical_factFINAL(spanish).pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/cervical_factFINALjan.08.pdf	View All Captures	Metadata
http://bcccp.ncdhhs.gov/linksandresources/EducationMaterialsandResources/fats.pdf	View All Captures	Metadata

Stop Words

- Application
- Memo
- Contract
- Poster
- FAQ
- Example
- Draft
- Worksheet
- Checklist
- Chart
- Instruction
- Deadline
- Invitation
- Schedule
- Flyer
- Survey
- Letter
- Program
- Rules
- Procedure
- Meeting
- Minutes
- Schedule
- Comment
- proposal
- Agenda
- Calendar
- Form
- Press release

Name	Size	Modified
..		11/26/2012 3:23:08 PM
reports.oah.state.nc.us.csv	4,460 kB	11/26/2012 3:21:43 PM
portal.ncdenr.org.csv	3,531 kB	11/26/2012 3:21:39 PM
www.ncdhhs.gov.csv	1,198 kB	11/26/2012 3:21:46 PM
www.ncagr.gov.csv	651 kB	11/26/2012 3:21:45 PM
www.ncdot.gov.csv	612 kB	11/26/2012 3:21:47 PM
www.ncleg.net.csv	491 kB	11/26/2012 3:21:48 PM
www.ncpublicschools.org.csv	473 kB	11/26/2012 3:21:48 PM
www.secretary.state.nc.us.csv	415 kB	11/26/2012 3:21:50 PM
info.dhhs.state.nc.us.csv	400 kB	11/26/2012 3:21:38 PM
www.dor.state.nc.us.csv	369 kB	11/26/2012 3:21:44 PM
www.nccommunitycolleges.edu.csv	320 kB	11/26/2012 3:21:45 PM
www.schs.state.nc.us.csv	281 kB	11/26/2012 3:21:49 PM
www.nccommerce.com.csv	277 kB	11/26/2012 3:21:49 PM
www.ncwater.org.csv	264 kB	11/26/2012 3:21:48 PM
www.osp.state.nc.us.csv	239 kB	11/26/2012 3:21:49 PM
www.nctreasurer.com.csv	193 kB	11/26/2012 3:21:50 PM
www.ncdoi.com.csv	150 kB	11/26/2012 3:21:46 PM
www.nchfa.com.csv	146 kB	11/26/2012 3:21:47 PM
www.jirdc.org.csv	125 kB	11/26/2012 3:21:44 PM
daq.state.nc.us.csv	121 kB	11/26/2012 3:21:38 PM
www.ncosc.net.csv	118 kB	11/26/2012 3:21:48 PM
www.ncforestservice.gov.csv	117 kB	11/26/2012 3:21:47 PM
www.ic.nc.gov.csv	117 kB	11/26/2012 3:21:44 PM
ncdhhs.gov.csv	113 kB	11/26/
www.hpo.ncdcr.gov.csv	107 kB	11/26/
www.ncdoj.gov.csv	100 kB	11/26/
www.deh.enr.state.nc.us.csv	99 kB	11/26/
www.nciom.org.csv	92 kB	11/26/
www.nccoastalmanagement.net.csv	87 kB	11/26/
epi.publichealth.nc.gov.csv	86 kB	11/26/
epi.ncpublichealth.com.csv	85 kB	11/26/
www.healthwellnc.com.csv	81 kB	11/26/
www.ncwildlife.org.csv	77 kB	11/26/
www.doc.state.nc.us.csv	76 kB	11/26/
www.nccourts.org.csv	74 kB	11/26/
www.eatsmartmovemorenc.com.csv	68 kB	11/26/
www.nutritionnc.com.csv	67 kB	11/26/
childnutrition.ncpublicschools.gov.csv	67 kB	11/26/
www.archives.ncdcr.gov.csv	53 kB	11/26/
www.governor.state.nc.us.csv	51 kB	11/26/
www.ncdps.gov.csv	51 kB	11/26/
www.correctionenterprises.com.csv	48 kB	11/26/
dcm2.enr.state.nc.us.csv	43 kB	11/26/
tobaccopreventionandcontrol.ncdhhs.gov.csv	43 kB	11/26/
www.ncmuseumofhistory.org.csv	39 kB	11/26/

	A
27318	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2001%20-%20general/30%20ncac%2001%20.0101.pdf
27319	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2001%20-%20general/30%20ncac%2001%20.0102.pdf
27320	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2001%20-%20general/30%20ncac%2001%20.0103.pdf
27321	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2001%20-%20general/30%20ncac%2001%20.0104.pdf
27322	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2001%20-%20general/30%20ncac%2001%20.0105.pdf
27323	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2002%20-%20%20covered%20persons%20and%20boards/30%20ncac%2002%20.
27324	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2002%20-%20%20covered%20persons%20and%20boards/30%20ncac%2002%20.
27325	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2002%20-%20%20covered%20persons%20and%20boards/30%20ncac%2002%20.
27326	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2003%20-%20ethics%20liasons/30%20ncac%2003.pdf
27327	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2004%20-%20agency%20heads/30%20ncac%2004.pdf
27328	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27329	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27330	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27331	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27332	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27333	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27334	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20a/30%2
27335	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20b/30%2
27336	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20b/30%2
27337	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20b/30%2
27338	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2005%20-%20statement%20of%20economic%20interest/subchapter%20b/30%2
27339	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2006%20-%20education/30%20ncac%2006.pdf
27340	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2007%20-%20gifts/subchapter%20a/30%20ncac%2007a%20.0101.pdf
27341	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2007%20-%20gifts/subchapter%20b/30%20ncac%2007b%20.0101.pdf
27342	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2007%20-%20gifts/subchapter%20b/30%20ncac%2007b%20.0102.pdf
27343	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2007%20-%20gifts/subchapter%20b/30%20ncac%2007b%20.0103.pdf
27344	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2007%20-%20gifts/subchapter%20c/30%20ncac%2007c%20.0101.pdf
27345	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2008%20-%20conflicts%20of%20interest%20and%20other%20ethical%20standar
27346	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2009%20-%20ethics%20complaints/subchapter%20a/30%20ncac%2009a%20.010
27347	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2009%20-%20ethics%20complaints/subchapter%20a/30%20ncac%2009a%20.010
27348	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2009%20-%20ethics%20complaints/subchapter%20b/30%20ncac%2009b%20.010
27349	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2009%20-%20ethics%20complaints/subchapter%20b/30%20ncac%2009b%20.010
27350	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20a/30%20ncac%2010a.pdf
27351	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20b/30%20ncac%2010b.pdf
27352	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20c/30%20ncac%2010c%20.0101.pdf
27353	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20c/30%20ncac%2010c%20.0102.pdf
27354	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20c/30%20ncac%2010c%20.0103.pdf
27355	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20c/30%20ncac%2010c%20.0200.pdf
27356	http://reports.oah.state.nc.us/ncac/title%2030%20-%20state%20ethics%20commission/chapter%2010%20-%20lobbying/subchapter%20c/30%20ncac%2010c%20.0301.pdf


User logs in
User uploads file list
CINCH locates URLs in file list
Actions CINCH performs on remote files
Duplicate check (previous download)
Allowable file extension?
File size assessed
X
Checksum calculated
X
Modified name assigned
CINCH downloads files
Actions CINCH performs on local files
Virus scan
X
Last modified date/time verified, reset
Metadata extracted
Checksum calculated
Duplicate check (current download)
AUDIT TRAIL for all actions
CINCH creates ZIP file(s)
CINCH emails user
User downloads ZIP file(s)
Repository
Bag-It
Manual review
KEY
X
If fails, file is deleted/not downloaded
If fails, file is added to problem_files folder

Problem files

Metadata, audit trail

Files

Name	Size	Type	Date Modified
problem_files		File Folder	6/19/2012 6:30 PM
event_list_2012_06_19.csv	31 KB	Microsoft Office Exc...	6/19/2012 6:30 PM
file_manifest_2012-06-19.csv	2 KB	Microsoft Office Exc...	6/19/2012 6:30 PM
PDF_Metadata_2012_06_19.csv	6 KB	Microsoft Office Exc...	6/19/2012 5:30 PM
statelibrary.ncdcr.gov_govdocs_classification_Classa.pdf	18 KB	Adobe Acrobat Doc...	5/27/2011 8:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classb2.PDF	21 KB	Adobe Acrobat Doc...	3/27/2007 2:09 PM
statelibrary.ncdcr.gov_govdocs_classification_classb.PDF	14 KB	Adobe Acrobat Doc...	4/26/2012 11:25 AM
statelibrary.ncdcr.gov_govdocs_classification_classc.PDF	37 KB	Adobe Acrobat Doc...	4/7/2008 10:45 AM
statelibrary.ncdcr.gov_govdocs_classification_classd2.PDF	60 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classd.PDF	97 KB	Adobe Acrobat Doc...	3/18/2004 9:44 AM
statelibrary.ncdcr.gov_govdocs_classification_classf.pdf	16 KB	Adobe Acrobat Doc...	5/27/2011 8:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classg2.PDF	22 KB	Adobe Acrobat Doc...	4/4/2006 11:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classg3.pdf	107 KB	Adobe Acrobat Doc...	2/7/2012 1:40 PM
statelibrary.ncdcr.gov_govdocs_classification_classg4.PDF	110 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classg.PDF	109 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classh.PDF	55 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi2.PDF	80 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi3.PDF	83 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi4.PDF	11 KB	Adobe Acrobat Doc...	7/6/2006 9:17 AM
statelibrary.ncdcr.gov_govdocs_classification_classi.PDF	26 KB	Adobe Acrobat Doc...	4/26/2012 11:25 AM
statelibrary.ncdcr.gov_govdocs_classification_classIntro.pdf	159 KB	Adobe Acrobat Doc...	3/18/2004 9:29 AM

Name	Size	Type	Date Modified
problem_files		File Folder	6/19/2012 6:30 PM
event_list_2012_06_19.csv	31 KB	Microsoft Office Exc...	6/19/2012 6:30 PM
file_manifest_2012-06-19.csv	2 KB	Microsoft Office Exc...	6/19/2012 6:30 PM
PDF_Metadata_2012_06_19.csv	6 KB	Microsoft Office Exc...	6/19/2012 5:30 PM
statelibrary.ncdcr.gov_govdocs_classification_Classa.pdf	18 KB	Adobe Acrobat Doc...	5/27/2011 8:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classb2.PDF	21 KB	Adobe Acrobat Doc...	3/27/2007 2:09 PM
statelibrary.ncdcr.gov_govdocs_classification_classb.PDF	14 KB	Adobe Acrobat Doc...	4/26/2012 11:25 AM
statelibrary.ncdcr.gov_govdocs_classification_classc.PDF	37 KB	Adobe Acrobat Doc...	4/7/2008 10:45 AM
statelibrary.ncdcr.gov_govdocs_classification_classd2.PDF	60 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classd.PDF	97 KB	Adobe Acrobat Doc...	3/18/2004 9:44 AM
statelibrary.ncdcr.gov_govdocs_classification_classf.pdf	16 KB	Adobe Acrobat Doc...	5/27/2011 8:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classg2.PDF	22 KB	Adobe Acrobat Doc...	4/4/2006 11:12 AM
statelibrary.ncdcr.gov_govdocs_classification_classg3.pdf	107 KB	Adobe Acrobat Doc...	2/7/2012 1:40 PM
statelibrary.ncdcr.gov_govdocs_classification_classg4.PDF	110 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classg.PDF	109 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classh.PDF	55 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi2.PDF	80 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi3.PDF	83 KB	Adobe Acrobat Doc...	3/18/2004 9:30 AM
statelibrary.ncdcr.gov_govdocs_classification_classi4.PDF	11 KB	Adobe Acrobat Doc...	7/6/2006 9:17 AM
statelibrary.ncdcr.gov_govdocs_classification_classi.PDF	26 KB	Adobe Acrobat Doc...	4/26/2012 11:25 AM
statelibrary.ncdcr.gov_govdocs_classification_classIntro.pdf	159 KB	Adobe Acrobat Doc...	3/18/2004 9:29 AM

URL

original file name

Metadata Field	File Location Derived From
Author	The file header - embedded metadata
Creation date and time	The file header - embedded metadata
Last modified date and time	The file header - embedded metadata
Creator	The file header - embedded metadata
Producer	The file header - embedded metadata
File name	The original URL, stripped of special characters. May include a unique number as well, to avoid duplication.
Title	The file header - embedded metadata
Number of pages	The file header - embedded metadata (a null return indicates the number of pages could not be determined)
Subject	The file header - embedded metadata
Keywords	The file header - embedded metadata
Licensed to	The file header - embedded metadata
Possible title	Best guess, extracted from full text (if any)
Possible keywords	Best guess, extracted from full text (if any)
Checksum (SHA-1)	Computed by CINCH
Whether or not full text exists	Determined by CINCH

CINCH: Capture INgest & CHecksum Tool

Home | About the Tool | About the Project | FAQ | Contact | Login

CINCH (Capture INgest CHecksum) is a tool that automates the transfer of online content to a repository, using ingest technologies appropriate for digital preservation.

CINCH
digital preservation . . . easier

How does CINCH work? Users:

1. Upload a list of URLs to files they want CINCH to retrieve.
2. Run the tool.
3. Download the retrieved files and file metadata to their local environment.

The tool will maintain original file timestamps; rename files with unique, intelligent identifiers; run virus scans; extract file-level metadata; create file checksums; validate checksums for file integrity; and track all actions in an audit trail document. With repeated uses, the CINCH tool will track file duplicates.

GET CINCH | USE CINCH | Documentation (.pdf)

Funding for the CINCH: Capture, Ingest, & Checksum tool is made possible through an IMLS Sparks! Ignition grant.

NORTH CAROLINA DEPARTMENT OF CULTURAL RESOURCES
WWW.NCCULTURE.COM

STATE LIBRARY of NORTH CAROLINA

NC live .org

INSTITUTE of Museum and Library SERVICES

Where can I get it?

slnc-dimp.github.com/Cinch/

I want more information!

cinch.nclive.org

Feedback welcome!

digital.info@ncdcr.gov

Funding for the CINCH: Capture, INgest, & CHecksum tool made possible through an IMLS Sparks! Ignition grant.