



# Pathways to Preservation

## Digital Curation Strategies in North Carolina State Government

North Carolina State Archives and State Library of North Carolina

April 2, 2009

DigCCur 2009





## Brief History of NC State Gov't Information Curation

---

- ❑ Two divisions within the Department of Cultural Resources
  - The State Library of North Carolina
    - ❑ Targets a variety of statewide needs (Library Development, LBPH, G&H)
    - ❑ The Government and Heritage Library provides direct services to state government employees and researchers
      - State Publications Clearinghouse collects and processes state agency created publications
      - Digital Information Management Program develops recommendations and tests solutions that support the identification, collection, cataloging, storage, and preservation of state government information for permanent public access
  - The North Carolina State Archives
    - ❑ Provides services to state government and public
      - Designated the official archival agency of the State of North Carolina;
      - Responsible for providing advice as well as records management program
      - Public records include all formats regardless of "physical form or characteristics"
      - Collects variety of formats of archival records
  - Separate mandates, missions, administrative structures
  - Mandates overlap, often work together to solve challenges
- ❑ Information Technology
  - DCR-IT provides desktop support
  - ITS provides network support
  - No defined support for long-term management of born-digital state records at this time



# Why is digital curation our problem?

- ❑ Statutory mandates
  - Gather, provide access to, and permanently store state agency electronic publications and records in North Carolina
- ❑ Trained professionals (Librarians & Archivists)
- ❑ Administrative support

|          | Pubs | Websites | Geospatial Data | E-mail | Databases | Other Records |
|----------|------|----------|-----------------|--------|-----------|---------------|
| Archives |      | √        | √               | √      | √         | √             |
| Library  | √    | √        |                 |        | √         |               |

*Challenges: Statutory mandates, but no enforcement authority and no programmatic IT support*



---

What we're doing today to meet the challenges of digital curation



## Websites: Harvests

---

- Archive-It (Internet Archive)
  - Pilot partners in 2005
  - Harvests began in fall 2005
  - Currently collected nearly 3TB
  - This spring: combining our harvests with relevant content harvested by Internet Archive between 1995 and 2005
- Challenges
  - Agency web design vs. tool limitations
  - Don't have 100% buy-in
  - Crawl analysis



# Websites: Access

- Access (short/long-term)
  - ARC/WARC file format
  - Annual backup to tape; stored locally
  - Multiple access points
    - NC State Gov. Web Site Archives, Webfeat search, Wayback machine, Archive-it page
- Anecdotal support
  - Governor Hunt speeches
  - Governor Easley executive orders

You are viewing an archived Web site, archived on 22/28/29 Nov 11, 2005, that is part of a collection of archived websites created using [Archive-It](#). The information on this web page may be out of date. External links, forms, and search boxes may not function within this collection. [\[hide\]](#)

**NORTH CAROLINA**  
Office of the Governor

Gov. Hunt Announces Framework To Attain "First In America" Goal  
Hurricane Relief Information  
Meet the Governor

Welcome from the Governor  
Governor Hunt's Biography  
Tour the Executive Mansion  
Meet the First Family

Report to the People  
The Hunt Administration  
1993 - 2001  
Record of Accomplishments  
negative Adobe Acrobat to view

James B. Hunt Jr.  
Governor

Initiatives & Accomplishments  
Learn more about Governor Hunt's initiatives, such as Smart Start, First In America 2010, the Excellent Schools Act and other administration accomplishments.

Governor's Office  
Visit the Governor's Office to learn of the many sections that keep the Governor's Office successfully operational.

Kids' Page  
Information for Kids

North Carolina's Initial First In America Report Card Video Conference  
(Real Player® is required.)

"Our greatest responsibility is to make sure that every North Carolinian gets a good start in life and a good education. For education not only enables us—it enables us."

CLEAN NC 2000  
NC SCHOOLS  
**FIRST IN AMERICA**  
2010

VISION 2030  
SAVING THE NEW ECONOMY  
GROWTH & OPPORTUNITY

Vision 2030 Project  
Non-profit & Volunteerism State Agency Liaisons

Home • Search



# State Publications

---

- Acquisition
  - Via e-mail, CD, or any other way we can get it
- Access & Storage:
  - Hosted access & storage – the OCLC suite of tools
    - Began using CONTENTdm (CDM), Connexion Digital Import (CDI), and the Digital Archive (DA) summer 2007
    - Current storage: ~1.5GB
    - CDI
      - the WorldCat record is automatically modified to include a link to the CDM record in the 856 field
      - a copy of the file is automatically passed to and stored in DA
      - the WorldCat record is automatically crosswalked from to DC and the file and the new DC record are added to CDM
        - DC metadata is manually edited to include additional info. (e.g., creating/submitting Agency)
    - Records findable via local website, search engines, Webfeat, WorldCat
  - Local storage
    - Local Storage Area Network (SAN)
    - Limited space, funds for growth



# Publications: Pre- and Post-Processing

|     |     |   |
|-----|-----|---|
| 040 |     | NCS #c NCS #d OCL #d NCS  |
| 029 | 1   | CNTNT #b /p249901coll22   |
| 037 |     | 02200 #b NCTE   |
| 043 |     | n-us-nc   |
| 086 |     | T3 7:A18 date #2 ncdocs   |
| 092 |     | #b  |
| 049 |     | NCSS  |
| 110 | 2   | North Carolina State Board of Certified Public Accountant Examiners.  |
| 245 | 1 0 | Activity review / #c North Carolina State Board of Certified Public Accountant Examiners.                                 |
| 260 |     | Raleigh, N.C. : #b The Board,   |
| 300 |     | v. ; #c 28 cm.  |
| 310 |     | Five issues yearly  |
| 362 | 0   | 1984, issue 4-  |
| 500 |     | Title from caption.   |
| 515 |     | Issue numbering begins each year with 1.  |
| 610 | 2 0 | North Carolina State Board of Certified Public Accountant Examiners #v Periodicals.                                       |
| 650 | 0   | Accountants #z North Carolina #v Periodicals.   |
| 780 | 0 0 | North Carolina State Board of Certified Public Accountant Examiners. # State Board activity review #w (OCoLC)19953193     |
| 856 | 4 1 | #u <a href="http://worldcat.org/oclc/20740890/viewonline">http://worldcat.org/oclc/20740890/viewonline</a> #3 View Online |

Connexion (MARC) view

1101 Oberlin Road, Suite 104 • PO Box 12827 • Raleigh, NC 27605 • 919-733-4222 • No. 7-2008

**Failure to Renew By Deadline May Result in Forfeiture**

21 NCAC 08J .0101, *Annual Renewal of Certificate, Forfeiture, and Reapplication*, requires that all active CPAs renew their certificates annually by the first day of July.

NOTE: Online license renewal will remain available through the end of

**Should You Register Your Firm With the Board?**

Although many North Carolina CPAs consider business, industry, government, or education as their primary employment, a significant number of these individuals also provide or offer to provide accounting services to the public.

A CPA who uses the CPA title in or

Statements on Standards for Accounting and Review Services; or

- any agreed-upon procedure or engagement to be performed in accordance with the Statements on Standards for Attestation Engagements.

Please note that a CPA or CPA firm that provides any audit services,

---

**Title** Activity review  
**Creator** North Carolina State Board of Certified Public Accountant Examiners.  
**Subjects** North Carolina State Board of Certified Public Accountant Examiners--Periodicals. Accountants--North Carolina--Periodicals.  
**Description** Monthly; Title from caption.; Issue numbering begins each year with 1; Electronic holdings begin 2000  
**Publisher** North Carolina State Board of Certified Public Accountant Examiners  
**Agency** N.C. State Board of Certified Public Accountant Examiners, Occupational Boards  
**Rights** State Document see <http://digitalstatelibnc.cdmhost.com/u/?p249901coll3,2>  
**Physical Characteristics** v. ; 28 cm.  
**Collection** North Carolina State Documents Collection. State Library of North Carolina  
**Type** Text  
**Language** eng  
**Format** Periodicals  
**Digital Collection** North Carolina Digital State Documents Collection  
**Digital Format** application/pdf  
**Related Items** <http://worldcat.org/oclc/20740890/viewonline>  
**Title Replaces** North Carolina State Board of Certified Public Accountant Examiners. State Board activity review  
**Audience** All  
**OCLC number** 20740890

**User-defined tags:**

- Statistics

Comments (0)

Join Mailing List Powered by CONTENTdm DIMP State Library of North Carolina Department of Cultural Re

CONTENTdm (Dublin Core) view



# Publications: Answering the Challenges

---

- Grants participation:
  - Washington State Digital Archives
    - NDIIPP grant through Oct. 2009, extension requested
    - Establish the Washington State Digital Archives (WSDA) model as one that can be replicated in other states and/or as a viable hosting service
    - No North Carolina publications submitted yet
  - iRODS & DCAPE (NHPRC grant through May 2011)
    - Build a distributed production preservation environment that meets the needs of archival repositories for trusted archival preservation services
    - Currently identifying required rule set for distributed data management environment and attempting to load Archive-It and CDM files



# State Records

---

- Acquisition
  - Via external media or hosted service
- Access & Storage:
  - Dark Storage— stored off-line or on external drives
  - Access to catalog via Manuscripts and Archives Reference System
    - Current storage: ~2 TB
    - Access provide upon request. No on-line access at the moment to the record.  
Access to the metadata
  - Local storage
    - Local Storage Area Network (SAN)
    - Limited space
    - Identifying funds for growth and to provide access
  - Pursuing Grants to build programs and test tools
  - Government Initiatives—Senate Bill 6.12



# Records: Answering the Challenges

---

## □ Grants participation:

### ■ GeoMapp

- NDIIPP grant through December 2009, extension requested
- To do a demonstration of transfer of archival geospatial data as well as test data from counties.
- Begin building capability within the State Archives to ingest and provide access to superseded data
- Moving into the phase of transferring data

### ■ iRODS & DCAPE (NHPRC grant through May 2011)

### ■ EMCAP (NHPRC grant through June 30, 2009)

- Development of email collection and preservation tools
- Builds on existing protocols and behaviors
- Keeps originals and transforms them to XML



# Geospatial Data--GeoMAPP

Table 3: Archival Activity by Data Type and Organization Type

| Counties (64)         |                  |                  |                |                  |
|-----------------------|------------------|------------------|----------------|------------------|
| Response              | Archiving Raster | Archiving Vector | Archiving Both | Archiving Either |
| Yes                   | 63%              | 73%              | 47%            | 89%              |
| No                    | 38%              | 27%              | 53%            | 11%              |
| Municipalities (40)   |                  |                  |                |                  |
| Response              | Archiving Raster | Archiving Vector | Archiving Both | Archiving Either |
| Yes                   | 65%              | 68%              | 48%            | 85%              |
| No                    | 33%              | 30%              | 50%            | 13%              |
| All Respondents (104) |                  |                  |                |                  |
| Response              | Archiving Raster | Archiving Vector | Archiving Both | Archiving Either |
| Yes                   | 63%              | 71%              | 47%            | 88%              |
| No                    | 36%              | 28%              | 52%            | 12%              |

Table 5: Archival Rates of 'Focus' Data Layers

| Layer                   | County | Municipality | Overall |
|-------------------------|--------|--------------|---------|
| Parcel                  | 63%    | 45%          | 56%     |
| Street Centerlines      | 52%    | 43%          | 48%     |
| Jurisdictional Boundary | 44%    | 60%          | 50%     |
| Zoning                  | 38%    | 53%          | 43%     |
| Orthos                  | 66%    | 60%          | 63%     |



# Geospatial Data

## GeoMAPP Infrastructure Inventory: Geospatial

### Storage

| Data Type                                | Storage Type (SAN/ Direct Attached (RAID)/ External drive) | If yes, Compression Technique (zip, rar, mrsid, etc) | Uncompressed (GB) |
|--|--|--|-------------------|
| Vector (NCOM prodN SQL Server DB)        | SAN and RAID   |  |                   |
| Vector (NCOM other prodN SQL Server DBs) | SAN and RAID   |  |                   |
| Vector (NCOM download - shapefile)       | HD   | ZIP  | 7.3               |
| Raster (1993 DOQQ)                       | HD   | JPEG   | 195               |
| Raster (1998 DOQQ)                       | HD   | MrSID  | 586               |
| Raster (DRG)                             | HD   | MrSID  | 29.3              |
| Raster (Landcover)                       | HD   |  |                   |
| Raster (Local Imagery)                   | HD   | MrSID  | 13035             |
| Available/ unallocated (free) storage    |  |  |                   |
| <b>Totals:</b>                           |  |  | <b>13,852.60</b>  |



# E-mail

## □ Governor Hunt E-mail server



## □ Governor Easley

|   | Notes   | Folders | Items | Emails processed | Attachments saved | Emails affected | Duplicates found | Unique Attachments |
|---|---|---------|-------|------------------|-------------------|-----------------|------------------|--------------------|
| Jan-June 2006   | located in governor0607 and governor1206  | 2       | 10693 | 10693            | 1340              | 1034            | 190              | 1156               |
| July-December 2006  | located in governor0607 and governor1206  | 76      | 13834 | 13833            | 2298              | 1785            | 167              | 2134               |
| Archive 07  | located in governor1207; includes entire 12 months; last 6 months of sent mail                      | 95      | 18973 | 18973            | 2105              | 1624            | 256              | 1850               |
| Jan-June 2007   | located in governor office email from Jan 07 to June 07; has an additional folder, Agency Responses | 4       | 6219  | 6218             | 953               | 761             | 84               | 869                |
| Jan07-June07  | located in in governor0607  | 3       | 4795  | 4794             | 527               | 448             | 31               | 496                |
|   |   |         | 54514 |                  |                   |                 |                  |                    |
| The main error is the length of the complete pathname - will not convert. |   |         |       |                  |                   |                 |                  |                    |



# E-mail--EMCAP

Status:  Read/Write  Read Only  Closed

| Unit Name                      | Contact Email                | WillProcess                         | Creation Date      | Message Count | Last Processed    | Processed Last Run | Earliest Message  | Latest Message   | Address     |
|--------------------------------|------------------------------|-------------------------------------|--------------------|---------------|-------------------|--------------------|-------------------|------------------|-------------|
|                                | kelley.mcfarland@ncmail.net  | <input checked="" type="checkbox"/> | 5/15/2008 10:14 AM | 0             | 2/13/2009 5:02 PM | 0                  |                   |                  | kelley.mc   |
| Community and Citizen Services | allison.stivender@ncmail.net | <input checked="" type="checkbox"/> | 5/15/2008 10:14 AM | 0             | 2/9/2009 9:02 PM  | 0                  |                   |                  | allison.sti |
|                                |                              |                                     |                    | 0             |                   | 0                  |                   |                  | claire.ent  |
|                                |                              |                                     |                    | 39071         | 2/19/2009 2:57 AM | 1075               | 1/29/2003 3:39 PM | 9/5/2033 4:32 PM | archivesg   |

### Reports Menu

Choose which accounts to include

- Selected account  
**archivesgovoffice@governor.emcap.ncdcr.gov**
- All accounts within the selected domain.  
**Citizen Affairs (governor.emcap.ncdcr.gov)**
- All accounts in all domains

Filter Options

Include:  All Accounts  Only Open Account

Include only those registered after:

Choose one of the following report type

Message Counts

Account Details:

Conversion Errors

Conversion Error Charact

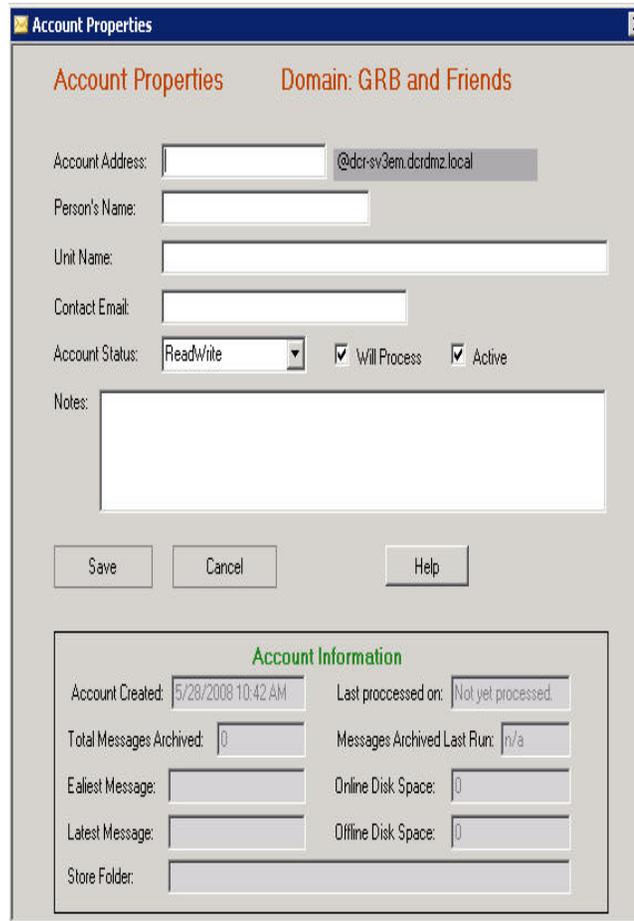
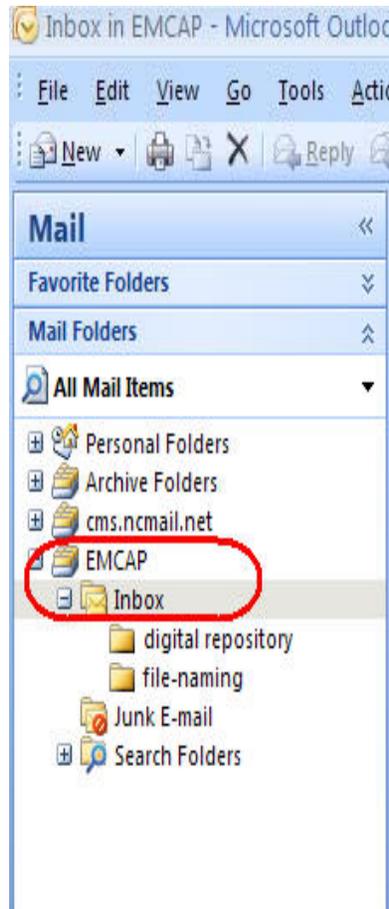
### Message Counts

3/24/200

| PersonName             | MessageCou                    | LastProcessed        | Status    | Account           |
|------------------------|-------------------------------|----------------------|-----------|-------------------|
| <b>Citizen Affairs</b> |                               |                      |           |                   |
| Agency:                | Community and Citizen Affairs |                      |           |                   |
| Domain Name:           | governor.emcap.ncdcr.gov      |                      |           |                   |
| <b>Accounts</b>        |                               |                      |           |                   |
| ArchivesGovOffice      | 39,071                        | 2/19/2009 2:57:38 AM | ReadWrite | archivesgovoffice |
| Total Message Count:   | 39,071                        |                      |           |                   |

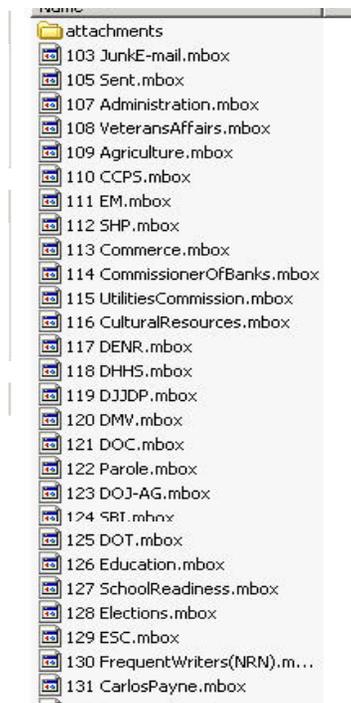


# EMCAP—client and administrative console





# EMCAP



```

</Folder>
<Folder>
  <Name>NCmail_Users</Name>
  <LocalId>3</LocalId>
  <Folder>
    <Name>NCmail User</Name>
    <LocalId>4</LocalId>
    <Message>
      <RelPath>attachments</RelPath>
      <LocalId>1</LocalId>
      <MessageId>&lt;000501c9b2e0$19e58360$4db08a20$@both@ncmail.net&gt;</MessageId>
      <MimeVersion>1.0</MimeVersion>
      <From>"David Both" &lt;david.both@ncmail.net&gt;</From>
      <To>"Kelly Eubank" &lt;kelly.eubank@ncmail.net&gt;</To>
      <To>"Sarah Joyner" &lt;sarah.joyner@ncmail.net&gt;</To>
      <Subject>Re: Fw: State policies on email as a public record</Subject>
      <Header>
        <Name>From</Name>
        <Value>"David Both" &lt;david.both@ncmail.net&gt;</Value>
      </Header>
      <Header>
        <Name>To</Name>
        <Value>"Kelly Eubank" &lt;kelly.eubank@ncmail.net&gt;; "Sarah Joyner"
      </Header>
      <Header>
        <Name>Subject</Name>
        <Value>Re: Fw: State policies on email as a public record</Value>
      </Header>
      <Header>
        <Name>Date</Name>
        <Value>Wed, 4 Dec 2002 09:53:31 -0400</Value>
      </Header>
      <Header>
        <Name>Message-ID</Name>
        <Value>&lt;000501c9b2e0$19e58360$4db08a20$@both@ncmail.net&gt;</Value>
      </Header>
      <Header>
        <Name>MIME-Version</Name>
        <Value>1.0</Value>
      </Header>
      <Header>
        <Name>Content-Type</Name>
        <Value>multipart/alternative; boundary="-----_NextPart_000_0001_01c9b2
      </Header>
      <Header>
        <Name>X-Mailer</Name>
        <Value>Microsoft Office Outlook 12.0</Value>
      </Header>
      <Header>
        <Name>X-01keid</Name>
        <Value>C8846F208EFD63EF6CAACA48AF27D9EFD03212B</Value>
      </Header>
      <MultiBody>
        <ContentType>multipart/alternative</ContentType>
    </Message>
  </Folder>
</Folder>

```

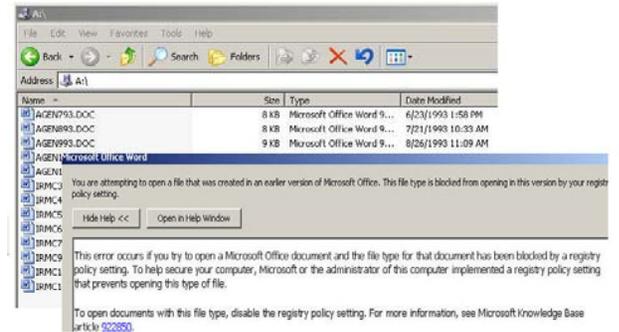


# Records

## IRMC-Information Resource Management Commission

### Governor Easley Files

- Paper and electronic
- 90.6 Gigabytes transferred
- Multiple file formats



### Files from the Press Office





# The Future

---

- ❑ Continue to monitor and test potential tools that would be a better fit or fill an outstanding need (like preserving databases)
- ❑ Continue drafting guidelines and “best practices” documents to provide guidance to state agency staff
  - Seek administrative approval
- ❑ Expand current education programs for state agency staff
  - their role in the process
  - importance of following guidelines to ensure that Library and Archives receive “preservation” ready digital info
- ❑ Work with IT at all levels to
  - Identify expectations & needs
  - Identify roles & responsibilities
  - Maximize staff & resources
  - Identify efficiencies



# Contact Information

---

Kelly Eubank  
North Carolina State Archives  
[kelly.eubank@ncdcr.gov](mailto:kelly.eubank@ncdcr.gov)

Jennifer Ricker  
State Library of North Carolina  
[jennifer.ricker@ncdcr.gov](mailto:jennifer.ricker@ncdcr.gov)

Amy Rudersdorf  
State Library of North Carolina  
[amy.rudersdorf@ncdcr.gov](mailto:amy.rudersdorf@ncdcr.gov)