Digital Preservation 101: What Libraries Should Know

INDIANA DIGITAL PRESERVATION

Justin Clark
Indiana Digital Preservation
Indipres.org

What is Indiana Memory?

- Launched in 2008
- Digital platform for Indiana's cultural heritage
- Collaboration with libraries and cultural heritage institutions
- Web portal to provide access

What is Indiana Memory?

- 150+ partners
- 400+ collections
- 600,000 items
- CONTENTdm Responsive
- Web Interface
- Integrated Features
- Working on harvesting from non-CONTENTdm

What is Hoosier State Chronicles?

- Launched in 2011
- Component of Chronicling America
- Digital platform for Indiana's historical newspapers
- Collaboration with libraries and cultural heritage institutions
- Web portal to provide access

What is Hoosier State Chronicles?

- www.hoosierstatechronicles.org
- Nearly 300 titles
- 155,000+ issues
- 1,500,000+ pages
- Text Searchable
- Downloadable (Both jpegs and PDFs)

Digital Preservation

❑ What is it?
  ❑ Defined as any planning and activity that helps ensure long term access to existing digital assets

❑ Why is it important?
  ❑ Hardware, software and operating system obsolescence
  ❑ Changes to file format
  ❑ Media degradation
  ❑ Threats to storage location
  ❑ Controlled access to assets
  ❑ LOCKSS principle

InDiPres

- ❏ Why InDiPres?
  - ❏ Trust
  - ❏ Community
  - ❏ Institutional Support

- ❏ Benefits
  - ❏ Creation of Digital Preservation Plan
  - ❏ Core Activity
  - ❏ Grants

InDiPres

- ❏ Offsite Dark Archive of Digital Files
    - ❏ Extra layer of security in the case of internal issue (fire, flooding, etc.)
- ❏ Decentralized Server Structure
    - ❏ Provides multiple backups in the event of a server failure
- ❏ Works with LOCKSS

INDIANA
DIGITAL
PRESERVATION

InDiPres

- ❑ 22 Members
  - ❑ Academic Libraries, Public Libraries, Historical Societies
- ❑ 14.9 TB of Archived Storage
  - ❑ Available to members via multiple notes in MetaArchive Network
- ❑ Works with a variety of CMS
  - ❑ Provides multiple backups in the event of a server failure

INDIANA DIGITAL PRESERVATION

Ingest Process

Select	Select content for preservation
Prepare	Prepare the content
Bag	Bag the content
Send	Send bags to the staging server
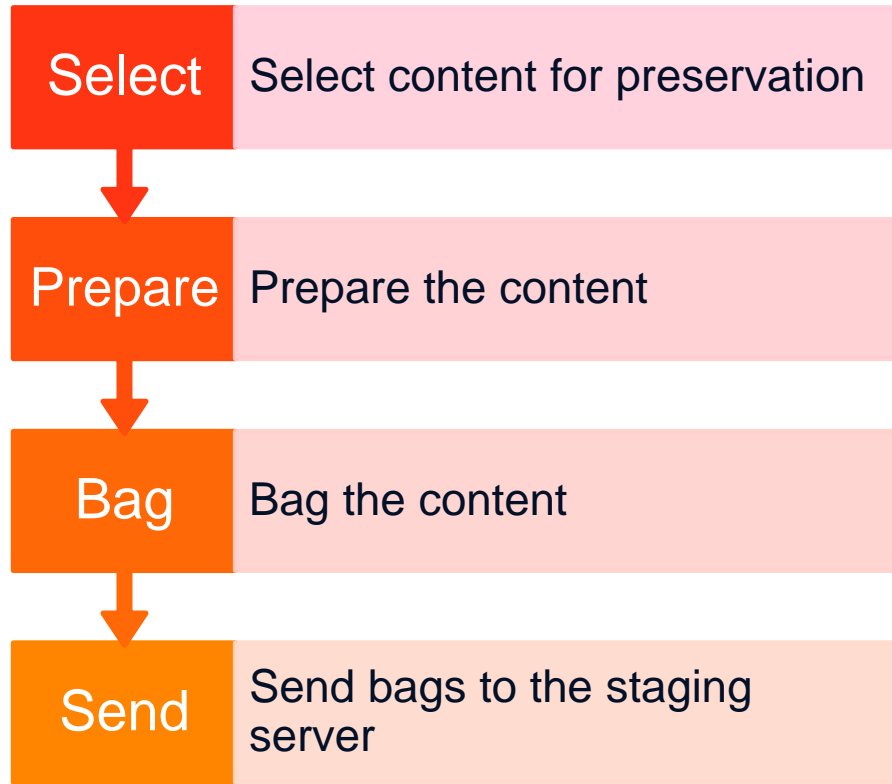
Ingest Process: Selecting Content

❑ Consider risk factors* such as vulnerability to change and loss, potential obsolescence, or disaster damage

❑ Prioritize* based on mission, policies, requirements, uniqueness, significance, frequency of use, age and extent of content, value, rights ownership, and digital quality

❑ *From the InDiPres Guidance Document and Technical Appendix

Ingest Process: Preparing Content

❑ Collect files
- ❑ Content files
- ❑ Metadata files
- ❑ Any other documentation that would help to put the collection back together again (field mapping information, etc.)

❑ Rename files, if needed
- ❑ Recommendations* include:
  - ❑ use prefixes that identify your institution and the collection
  - ❑ avoid spaces and special characters

❑ Use standard file formats (preferably non-proprietary)
- ❑ This will help for future format migrations

*See the InDiPres Guidance Document and Technical Appendix

Ingest Process: Bagging Content

❑ A bag* is like a folder or directory on a computer and is composed of:
- ❑ a bag declaration text file, which is like a seal of authenticity
- ❑ a text-file manifest listing the files in the collection
- ❑ a subdirectory - usually titled "data" – filled with the digital content
- ❑ an optional text file, titled "bag-info.txt," with administrative metadata

❑ We are currently using the Library of Congress Bagger software. There are other bagging software options, such as DART, from APTrust

- ❑ Help is available (documentation, training videos and sessions)

❑ Bag names should start with a prefix that identifies your institution and the collection and should include the bag number.

❑ *From the Library of Congress Digital Preservation page on Digital Content Transfer Tools

Ingest Process: Staging Content

❑ Ideally, the bags are sent to the staging server electronically, but it is also possible to deliver them on hard drives.

❑ When you have sent bags, please let your InDiPres contact know. It is also helpful if you include the bag names and sizes.

INDIANA
DIGITAL
PRESERVATION

Ingest Process: Testing and Ingest

❑ InDiPres will create Archival Units (AUs) identifying the location of the bags on the staging server

❑ The MetaArchive test server will pull the bags and run tests

❑ MetaArchive LOCKSS server administrators at 5 geographically diverse locations will be notified to pull the bags and ingest them into their LOCKSS servers

❑ When the ingest is complete, the bags will be removed from the staging server to make room for more bags

INDIANA DIGITAL PRESERVATION

Current Ingest Process
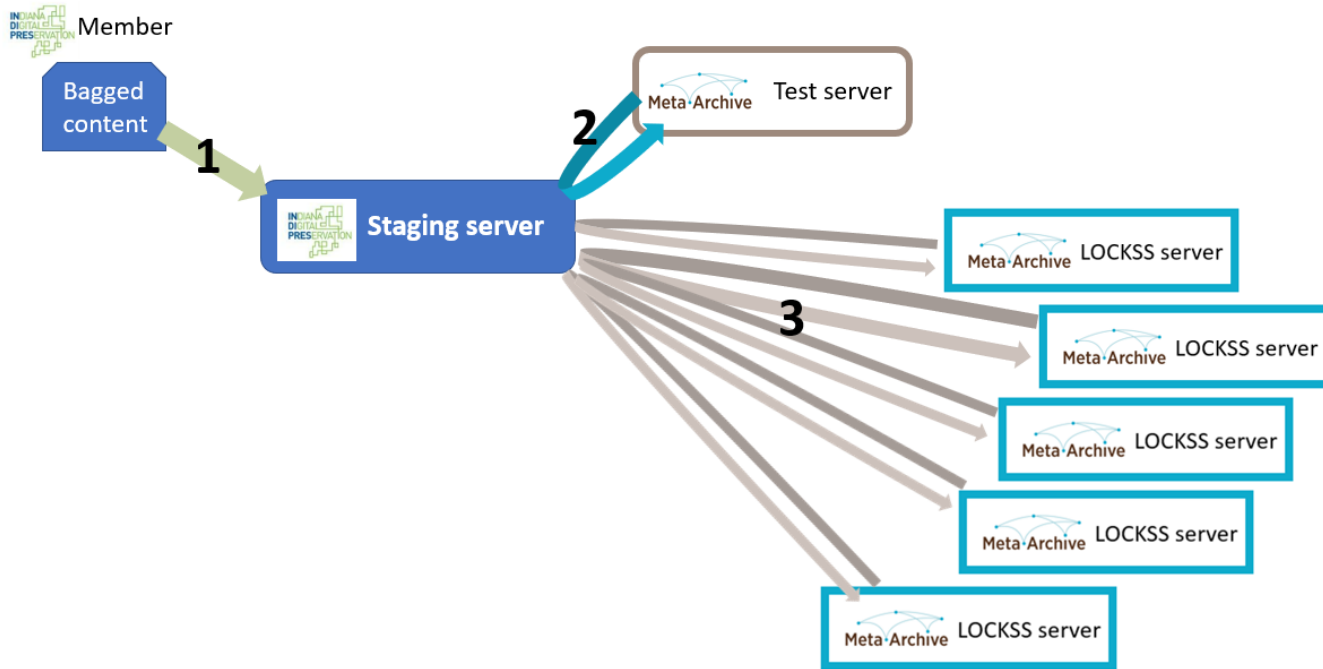
For More Information

MetaArchive: https://metaarchive.org/

InDiPres: https://indipres.org/ (see the Guidance Document under Member Tools)

Please feel free to contact us!

❑ Justin Clark: JusClark@library.IN.gov

❑ Charla Gilbert: cgilbert@vinu.edu

❑ Natasha Hollenbach: nhollenbach@indypl.org

INDIANA DIGITAL PRESERVATION

Thanks!

Questions?