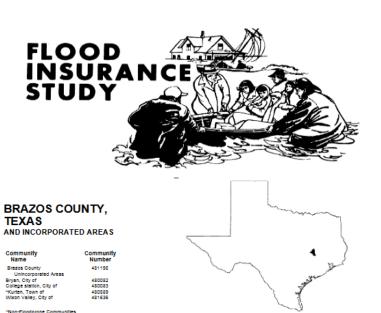
# The Government Documents Digitization Initiative: Shepherding Resources from Shelf to Server

Ryan Laddusaw Sean Buckner Laura Sare



# Flood Insurance Study

A Flood Insurance Survey (FIS) is a compilation and presentation of flood risk data for specific watercourses, lakes, and coastal flood hazard areas within a community.







# HathiTrust U.S. Federal Government Documents Program

### Mission:

Through coordinated and collective action, expand and enhance digital access to U.S. federal publications including those issued by GPO and other federal agencies.

### Goals:

Comprehensive digital collection: Build a comprehensive HathiTrust digital collection of federal documents distributed in print format by the Government Publishing Office as part of the Federal Depository Library Program.

Enduring access: Ensure that the benefits of HathiTrust work in areas such as improved quality, shared print collection management, collection development, enhanced resource discovery, and broader access to scholarly materials are realized in the HathiTrust federal documents collection and services.

https://www.hathitrust.org/usgovdocs



### From Shelf to Server

### Departments

- Evans Subject Librarians Government Information
- Metadata & Cataloging serials cataloging/metadata
- Collection Development & Acquisitions GovDocs
- Preservation Digitization and Digital Preservation
- Digital Initiatives scripts, EZID and HathiTrust

### Migration of Content

- Interdepartmental processes involved
- Multiple metadata transformations
- Digitized items renamed multiple times
- Files move across 4+ servers/spaces, various copies and formats generated
- Physical items: only TX records retained







### Workflow Development

- Pull documents from shelves who, how many, when?
- Check records OCLC, upgrade/create records, add ARK/URL, update ILS TX vs. non-TX, generate metadata records XML/spreadsheet
- Digitize disbind, scan, process TIFFs, file renaming, data transfer/file management; digital preservation?
- Gov Docs rebind or recycle, records management
- Scripts generate ARKs, build SIPs: derivative files-attach metadata, upload metadata to EZID-publish ARKs



# Archival Resource Keys

- An ARK is a URL that is a persistent identifier
- Randomly generated alphanumeric string (e.g. ark:/81423/m3a1b2) appended to a handle
- TAMU Libraries uses EZID service to create and maintain ARKs
- Required by HathiTrust for content submissions
- Must use <u>Pairtree</u> file system convention for SIPs ark:/81423/m3a1b2 → ark+=81423=m3a1b2
- Preservation requested a prefix be added for archival/ storage copies: gddi\_0001\_ark+=81423=m3a1b2



## Workflow

Generate JPEG2000s

OCR (Tesseract)

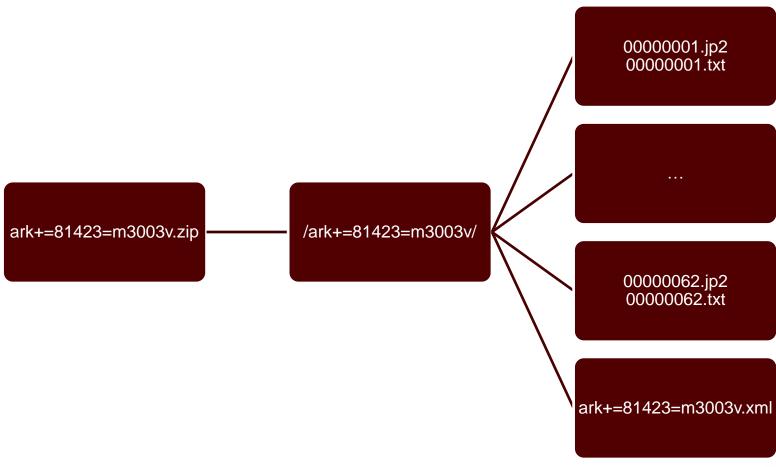
Create XML Metadata

Package files for ingest

Submit to HathiTrust

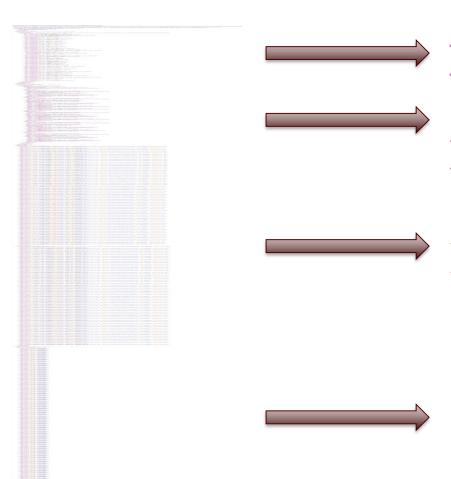


# Structure





# **METS XML Metadata**



### MARC Record

### PREMIS Metadata

<PREMIS: eventIdentifier>
 <PREMIS: eventIdentifierType>UUID</PREMIS: eventIdentifierType>
 <PREMIS: eventIdentifierValue>1944ee31-76f9-359f-baa6-b30df6599cfe</PREMIS: eventIdentifierValue>
 </PREMIS: eventIdentifier>

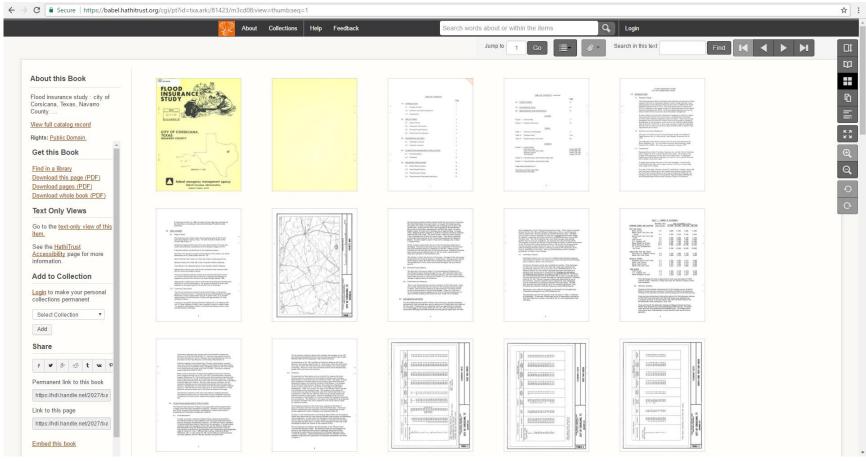
### **Checksum Data**

### Structural Map

<METS:div ORDER="5" TYPE="page">
 <METS:fptr FILEID="0CR00000005"/>
 <METS:fptr FILEID="IMG00000005"/>
 </METS:div>

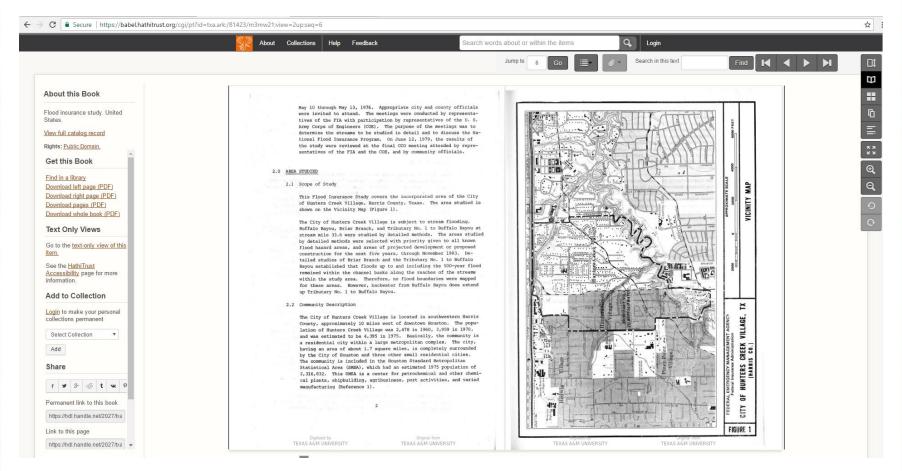


# Thumbnail View



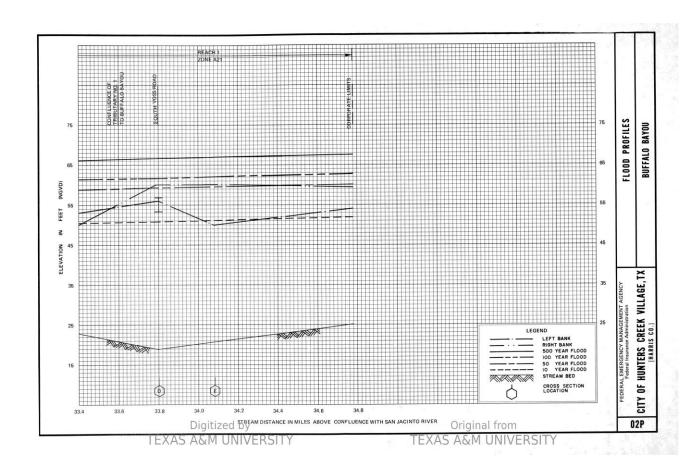


# Page-Turner View





# Chart





# Chart - Detail







Ryan Laddusaw – rladdusaw@tamu.edu Sean Buckner – sbuckner@tamu.edu Laura Sare – lsare@tamu.edu

TEXAS A&M