

Archives as Data 2024

Archiving Digital Records Workshop

Thanks to our

- sponsor:
 - NEH Institutes for Advanced Topics in the Digital Humanities
- organizations:
 - Columbia Libraries
 - History Lab

Orientation

- wifi: Columbia University
- restrooms
- location
 - Lehman Suite, 406 International Affairs Building
 - International Affairs Building
 - Morningside Campus

Daily Schedule (Week 1)

- 09:00 AM: Meet in front of the gates at 116th and Broadway
- 09:30 AM-11:55 AM: Technical Lecture & Exercises
- 12:00 PM-01:00 PM: Lunch & Speaker
- 01:00 PM-02:00 PM: Discussion with speaker
- 02:00 PM-03:00 PM: Lab or reading discussion

Daily Schedule (Week 2 - Remote)

- 12:00 PM-01:00 PM: Lunch & Speaker
- 01:00 PM-02:00 PM: Discussion with speaker

Introductions

- Name
- Role
- Affiliation
- Motivation/Goal

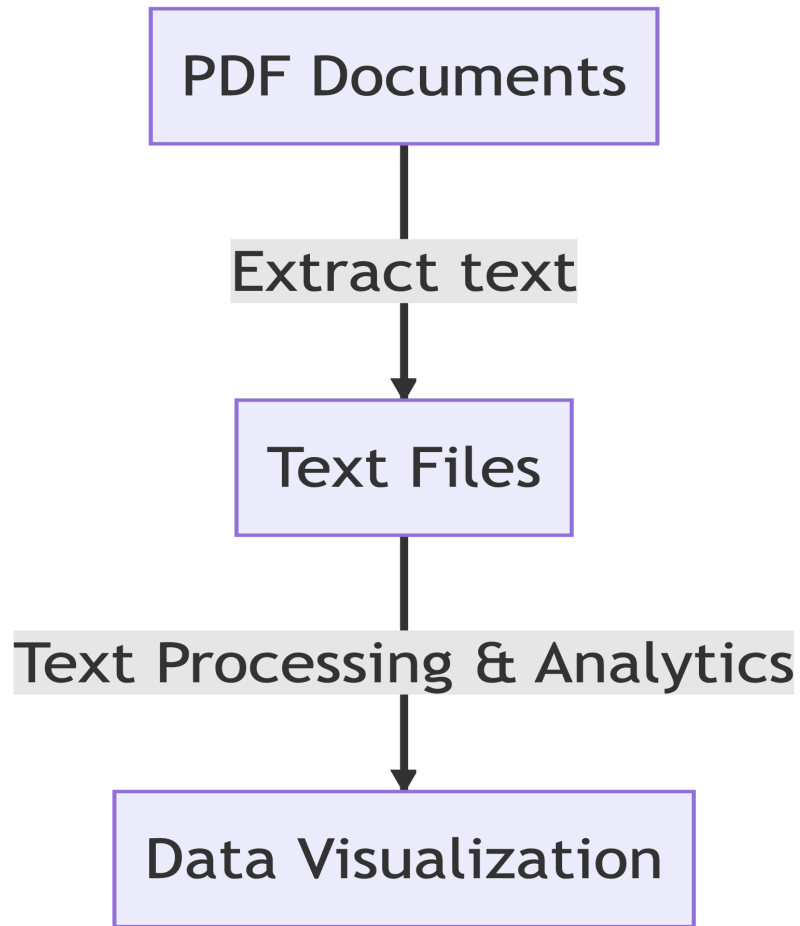
Lightning Talks

- You speak on a topic of interest
- ~5 minutes
- Tuesday @ 2:00 PM
- Optional but encouraged

Objectives for Technical Sessions

1. Gain PDF knowledge
2. Get comfortable with the command line
3. Introduced to text analytics tools
4. Extract text from PDFs and apply text analytics
5. Become a more effective team member on DH/IT projects
6. Be more willing to try DIY projects
7. Gain awareness of shortcomings & pitfalls

Target Processing Workflow



Points to keep in mind

- Second time giving this workshop
- Active, point-in-time participation welcome
- The firehose
- Ben's a technical person, not an archivist or historian