

Framework for Mindfully Planning Newspaper Digitization in Regards to Computational Analysis

Purpose: parallel to earlier work where we developed a quick description/overview on how UF did digital preservation, and then we included that information in grants and we shared it out to explain our processes.

Goals of the 5/18 meeting include:

- Make sure we are all aware of this conversation, for use in thinking/working with for current and future grants and any other digital newspaper projects
- Identify folks who are interested in reviewing and contributing to the checklist/framework (all on 5/18 meeting)
- Confirm/update on needs/expectations for any technical concerns related to new UFDC for the patron interface and production workflows

For this work, we want to eventually have:

- Outputs:
 - Checklist/framework for planning digitization projects – planning document for digitization projects to build capacity and support computational analysis
 - Clarity on existing (and expected new) options for people (contact dLOC technical director or dLOC digital scholarship director; know needs to update DSS to support bulk downloads for text and mets/mods; like old grab text, METS, JPG and how this functionality needs to be part of rebuilding prod in 2022)
 - Documentation, building capacity, for use in the dLOC as Data report (and others as applicable)
- Outcomes:
 - Initial checklist/framework and discussions as a foundation for supporting future capacity development - see below
 - Support for how do we build our PIs for understanding on these needs, and for liaising for scholars
 - Add feasibility meeting for grants, for DH/data work
 - Add checkbox on grants checklists for digital+ like these computational considerations
 - Informs how we building new UFDC and rebuild production to support these needs
 - Ensure all files available under downloads?
 - Ensure anything produced with this are shared to AskA and library-wide

- Use cases on projects and uses - Perry will provide several, and we will build from, for considering changes to UFDC
- Consider building into grants to have easy-download of data as part of it, for specific, scoped sets for materials that are complete (e.g., full newspaper run) or for sets that are comprehensible
- Create LibGuide for how to play with these downloadable data sets, with resources - will have one from dLOC as Data

Checklist: Planning Newspaper Digitization Projects

1. Project/grant planning
 - a. Include an additional feasibility meeting for considerations of computational support
 - b. Note access to text (and other, metadata, etc.) files as added value in grant proposals, in context of specific use cases
 - c. Consider template language
 - i. Contact Perry Collins for use language on copyright
 - ii. Contact Laurie Taylor for general language on process
2. Check in with technical contacts for project planning
 - a. Project-specific: NDNF Coordinator, dLOC Technical Director
 - b. General for projects: Digital Partnerships & Strategies (Copyright & OER Librarian and Chair), DSS Chair, LTS Chair
 - c. Liaison-area specialist: DH Specialist, ARCS, Research Computing
3. Review existing resources/options for text analysis, which include:
 - a. Utilize Voyant Tools and/or Palladio
 - b. Contact ARCS for support/expertise in text mining and analysis
 - c. Contact Research Computing for support/expertise in text mining and analysis
 - d. Review existing supports and data sets for a specific project, as with <https://dlocasdata.domains.uflib.ufl.edu/>
4. Review existing resources/options for gathering text files:
 - a. Text files are available in UFDC
 - b. Utilize new UFDC ElasticSearch for searching
 - c. Contact DSS for bulk set of files (txt, METS/MODS, PDF)
 - d. Use browser plugin DownloadThemAll for bulk download
 - e. Utilize the new UFDC APIs for file download
 - f. Contact LT&DS for liaising on ElasticSearch functionality, to learn what could be considered for native search in new UFDC
 - g. Consider project/program-specific additional functionality, as with
 - i. Chronicling America (APIs and bulk downloads for data): <https://chroniclingamerica.loc.gov/about/api/>
 - ii. dLOC as Data: <https://dataverse.fiu.edu/dataverse/dloc-as-data>
 - iii. Specific to paid databases - contact your liaison librarian
 - iv. Contact technical folks for planning specific, scoped downloads for getting started