

LAMP Spring 2022 Digitization Proposal
Reeling in the French Antilles: digitizing newspapers from Martinique and Guadeloupe, 1852 to 1929
University of Florida

I. Narrative

Abstract

Reeling in the French Antilles: digitizing newspapers from Martinique and Guadeloupe, 1852 to 1929

The George A. Smathers Libraries at the University of Florida (UF) requests [see Cost Summary] to support reel preparation activities and vended digitization costs related to digitizing 31,200 pages of UF microfilm holdings of French-language newspapers published in Martinique and Guadeloupe between 1852 and 1929. Titles selected for digitization include *Le Propagateur* (Saint Pierre, Martinique) and *Journal officiel de la Guadeloupe* (Basse-Terre, Guadeloupe). Duplicates of archival master microfilm from UF will be used to complete the digitization. Additionally, these duplicates will be added to the UF Latin American & Caribbean Collection Reading Room as access/use copies as an additional access point. The digitized content will be freely available to the public through the Digital Library of the Caribbean (dLOC, <http://dloc.com>) and the Caribbean Newspaper Digital Library (CNDL), and copies of the digital content will be provided to CRL.

Content

Martinique and Guadeloupe were colonized by France in the mid-17th century, though there were periods of time when these islands, also known as the French Antilles, were governed by other foreign powers. The early presses in the Caribbean, especially in the French Antilles, were known for reprinting colonial news, a way to keep colonists connected to their motherlands. Correlations can be drawn from events related to slavery and trade/commerce to freedom of the press. Slavery was abolished in the French colonies in 1848, the same year that a decree denounced censorship of the press, allowing newspapers the freedom to print content without authorization from the government and protecting them from being suspended or revoked. As a result, newspapers were not allowed to promote the re-establishment of slavery, however, on the same token were not allowed to insult government officials or to incite resistance to the French government.

Between 1852 and 1929, the press in French colonies was primarily made up of official notices and provided a glimpse into the life of the time on the islands through reporting of general and local news. We propose to digitize two of these newspapers from the French Antilles: 1) *Le Propagateur*, published in Saint-Pierre, the cultural capital of Martinique and 2) the *Journal Officiel de la Guadeloupe*, published in Basse-Terre, the administrative capital of Guadeloupe. *Le Propagateur*, was founded in 1854 and published weekly until 1894. The *Journal* began publication in 1815 as the *Gazette Officiel*, a newspaper established by the Governor to print laws and ordinances. The paper changed names to the *Journal Officiel* in January of 1882 and became the *Journal Officiel de la Guadeloupe* in July of that same year.

The University of Florida (UF) Libraries filmed these newspapers in 1951, prior to the establishment of the Preservation Department in 1987. The UF microfilm holdings have not been digitized previously. The *Propagateur* has not been digitized/made available elsewhere, however there are some issues for the *Journal* available online. There are currently 413 issues of the *Journal* from issues published in 1882, 1896, 1902-03, 1905 and 1936 available online to the public via [Gallica](#), the Bibliothèque nationale de France (BNF) digital library. The content UF has on microfilm that would be digitized for

this project would complement the digitized archives in BNF's digital library. Additionally, there are digital issues available online through [CRL's Digital Delivery System](#) for 1891, however these are another title, the *Bulletin Officiel de la Guadeloupe*, a periodical published by the French government in Guadeloupe which began printing in 1804 under the title *Journal officiel de la Guadeloupe et Dépendances*.

The University of Florida holds first-generation microfilm on acetate film for these titles. The master microfilm is stored with an off-site vendor in cold storage. There are a total of 26 microfilm reels (35 mm black and white, silver halide film) for these titles. We estimate that each reel contains 600 two-up frames, although the exact number of frames, pages, and issues per reel varies. There are six reels for *Le Propagateur* that contain approximately 7,200 pages for the years 1855 and 1893. During this time, the *Propagateur* published weekly. We have 20 reels for the *Journal*, which contain approximately 24,000 pages published between 1852 and 1929. The *Journal* was known to publish bi-weekly (twice a week) during this time, with each issue typically consisting of four, six, or eight pages. Appendix 1 includes UF's master microfilm inventory.

Note: Narrative sections below provide standard information for proposals submitted by the University of Florida (UF) George A. Smathers Libraries, when applicable, for all digital projects. This information is current as of March 2021.

Copyright / Permissions

Due diligence has been undertaken by UF to identify relevant source law. Per French law (followed in Martinique and Guadeloupe), copyright law has strong emphasis on moral rights. In accordance with moral rights, these works will not be altered in any way that could harm the creators or descendants and metadata standards will be followed to include attribution. Additionally, copyright duration in collective works such as a newspaper is 70 years and as such, these works are now in the public domain.

The Smathers Libraries support US Copyright Law as well as moral and cultural heritage rights, and other applicable rights. In order to support these rights for UF, partners, and constituencies, the Libraries follow a [permissions-based model](#). Full documentation on rights and permissions in place are maintained for all materials. If the permissions and rights in place allow the assignment of rights to LAMP, then those can be assigned. Additional information will be provided based on the specific project needs.

Open Access Commitment Outlined

The [Digital Library of the Caribbean](#) is a cooperative Open Access digital library for resources from and about the Caribbean and circum-Caribbean that provides access to digitized versions of Caribbean cultural, historical and research materials currently held in archives, libraries, and private collections. With 40 content contributing partners and sixteen financially supporting members, dLOC is a stable digital resource for Caribbean researchers worldwide. The [Caribbean Newspaper Digital Library](#) is a collaborative international digital newspaper library held within dLOC that preserves and provides access to valuable resources for the study of the Caribbean and the advancement of Caribbean Studies. CNDL includes 2,120,353 pages, in 237,278 items (see <https://original-ufdc.uflib.ufl.edu/cndl/itemcount> and <https://dloc.com/collections/cndl>).

From January to December 2021, the collection had 2,632,901 views (see <http://ufdc.ufl.edu/stats>).

Conversion Procedure

The digitization and OCR text-conversion will be executed to Library of Congress' National Digital Newspaper Program (NDNP) specification via contract with a qualified outside vendor. Resulting digital files will be made available through dLOC and CDNL, all hosted by UF as part of the University of Florida Libraries Digital Collections (UFDC).

What quality control will be used to ensure best practices are adhered to throughout the conversion process? UF expects vendors to work at or above industry standards for digitization and adhere to digital preservation standards. UF will conduct post-digitization inspections on all files to ensure that they meet contractual obligations.

If OCR is generated, will it be edited or uncorrected? OCR text is uncorrected.

Metadata

Metadata processing is common for all materials. [Metadata Encoding and Transmission Standard \(METS\)](#) metadata is created using the SobekCM tools and system, which power digitization and digital ingest production. As items are processed, the metadata is enhanced automatically and manually as objects move through the imaging/curation workflows. The system assigns a unique Bibliographic Identifier (BibID) to each object processed, and that BibID is used to track the item. The METS files include technical and structural data about each image, as well as descriptive and administrative information.

Any pre-existing metadata (e.g., from catalog records, finding aids, museum accession records) will be imported into the system at the first stage, before the start of imaging. The metadata for materials is prepared by Catalogers, Archivists, Subject Matter Experts, Registrars, Curators, and others as appropriate for the project.

UF's systems store all metadata in METS/MODS as well as automatically transforming and providing the metadata in MARCXML, with all metadata accessible online in the public access systems (e.g., UF Digital Collections, Digital Library of the Caribbean, and Florida Digital Newspaper Library). All materials are optimized for search engine access to ensure worldwide reach through Google and other search engines. The public access systems have integrated support for OAI-PMH ([Open Archives Initiative or OAI](#)) coming in Q1 of 2022 to ensure all metadata is harvestable following OAI-PMH standards.

Added-Value Features

Will text files be made searchable via the application of Optical Character Recognition software or double-keying? The public access systems (e.g., UF Digital Collections, Digital Library of the Caribbean, and Florida Digital Newspaper Library) provide full text searching within collections as well as having the collections and materials optimized for search engine access to ensure worldwide reach through Google and other search engines.

Will searchable text files be marked up in accordance with specific schema? No.

Will numerical files be rendered in forms suitable for statistical manipulation? Files are rendered by file type, with most statistical files as XLSX or similar, which are rendered as downloadable.

Will cartographic and related materials include geospatial referencing? Yes, when created for materials, the systems support inclusion of this in metadata and this powers display/filtering.

Access

Delivery system for this project will be the public access systems (e.g., [UF Digital Collections](#) and the [Digital Library of the Caribbean](#))

In what format will the files be delivered? Imaged object files are delivered online in single-page JPG, JPG2000, and JPG thumbnail images along with the OCR text files (TXT) and with the metadata, displayed as a “citation” and available and displayed in all metadata formats (METS/MODS, MARCXML).

Will the data be freely available on the internet? If not, what limitations to access will be in place for this data (and why)? All data will be freely available.

What search and browse capabilities will be used to access the data?

The public access system supports all collections and items including:

- Full text searchable
- Browseable - with browse views by title and thumbnail
- Serve text, image, multimedia, audio, video files, data sets, and more within the same collection
- Support for multiple file types (text, image, oversized images, video, audio)
- Powered by rich metadata support
- Google-map based searching or map browsing

Custom views for specific item-types include:

- Full-screen page turner view
- Sanborn maps
- Image zoom and pan viewing capabilities

Will the metadata allow for easy harvesting of data? Yes. All materials are optimized for search engine access (SEO) to ensure worldwide reach through Google and other search engines. Integrated support for OAI-PMH ([Open Archives Initiative or OAI](#)) to ensure all metadata is harvestable following OAI-PMH standards will be in place by end of Q1 of 2022.

Archiving

What is the process for sustained preservation of the files? Will the data be archived at any location(s) other than CRL? The Libraries are committed to long-term digital preservation of all materials in the UF Digital Collections and in UF-supported collaborative projects including Digital Library of the Caribbean (dLOC) and the Caribbean Digital Newspaper Library (CDNL). Redundant digital archives, adherence to proven standards, and rigorous quality control methods protect digital objects. The UF Digital Collections provide a comprehensive approach to digital preservation, including technical supports, reference services for both online and offline archived files, and support services by

providing training and consultation for digitization standards for long-term digital preservation.

The Libraries support locally created digital resources, including the UF Digital Collections which contains over 500,000 digital objects with over 14 million pages (as of January 2022). The Libraries create METS/MODS metadata for all materials. Citation information for each digital object is also automatically transformed into MARCXML and Dublin Core. These records are widely distributed through library networks and through search engine optimization to ensure broad public access to all online materials.

The Libraries maintain redundant servers with copies of all online and dark files on separate Library and University maintained servers, with an additional Tivoli tape backup as a ready-access archive. The tapes are maintained in an environmentally controlled storage facility. The servers and tape storage are in geographically separate locations. In practice consistent for all University of Florida Digital Collections and projects, separate redundant digital archives are maintained in UFAR. This infrastructure meets or exceeds [National Digital Stewardship Alliance \(NDSA\) levels 2 and 3](#) for digital preservation. Information about the archival processing for all digital objects, both online and offline, or for “dark” archived objects, is tracked and maintained within the current SobekCM online system under “Work History.” The SobekCM system is in the process of being replaced with the development of a new digital library user and internal system. During the development of the new system (expected 2020-2023), SobekCM will remain in use, and preservation information will be visible under “Work History” tracking including the “History” which lists the workflow, date the workflow occurred, and location/notes. Under “Work History,” another field titled “Archives” will lists all of the archived files including filename, size, last write date, and archived date.

How will you deliver the files to CRL? Files to partners are regularly transferred using FTP or mailed external hard drives, with both supported and selectable by partners for best applicability for their processing.

What will you do with the original source material? Decisions on the disposition of source material are handled by the appropriate collection manager, curator, or archivist. However, as UF currently does not have access copies for these titles, UF expects to add the second-generation access microfilm copies to the Latin American and Caribbean Collection Reading Room. The master film will be retained in cold storage to serve as a preservation master copy and allow to produce duplicate microfilm copies as needed.

II. Plan of Work

UF plans to work with Creekside Digital and Digital Divide Data to undertake the project. As with all vended digitization, UF will ensure quality control of metadata and materials, and open access and archiving of all materials from the dedicated webpages in dLOC for these titles.

The stages of this project consist of:

- Regeneration of microfilm
- Microfilm review, issue level report preparation, and shipping
- Vendor digitization
- Ingest of vended materials, which are digitized according to NDNP Specification
- TIFFs and single-page PDFs will be delivered to CRL via Dropbox, FTP, or other method.

The current work plan is designed to complete the project in one funding cycle.

Q1: regeneration of microfilm; the microfilm will be reviewed, an issue level page count and condition report will be prepared, and the reels shipped to digitization vendor. All shipping costs will be covered by UF.

Q2-4: Depending on vendor workflows after the funds are awarded, the vendor will convert the 26 reels in several batches scheduled across Q2-4. The vendor digitization and UF validation/ingest processes will be done for each of the batches and details on these activities are below. TIFFs and single-page PDFs will be delivered to CRL via Dropbox, FTP, or other method to be determined. Digital content will be ingested into appropriate aggregations within UFDC/dLOC.

In accordance with the NDNP specification, the vendor will scan from the microfilm and create derivative files according to specifications described in [Appendix B \(page 35\) of the National Digital Newspaper Program Technical Guidelines](#). The film will be scanned in 8-bit grayscale with a maximum resolution between 300-400 dpi, relative to the physical dimensions of the original material. Scanning will produce an uncompressed, unprocessed TIFF 6.0 file for each newspaper page on the microfilm. In cases where newspaper titles were microfilmed with two pages per frame, the vendor will adjust produce a single image file for each newspaper page. The scanned TIFF will be de-skewed and cropped to the page edge, if necessary. Prior to digitization of each reel, the vendor will scan a target. The vendor will scan a second target during the reel digitization to aid in monitoring of scan quality. Each target will be described appropriately in reel metadata

In addition to a TIFF 6.0 file for each newspaper page on microfilm, the vendor will produce a:

- JPEG 2000 file from the TIFF 6.0 file. The JPEG2000 file will conform to the 21 specifications listed in Appendix B of the NDNP Technical Guidelines. For instance, each JPEG 2000 file will have 6 decomposition levels, 25 quality levels and a compression ratio of 8:1.
- PDF file from the TIFF 6.0 file. The PDF file will conform to the 18 specifications listed in Appendix B of the NDNP Technical Guidelines. It will have a file name corresponding to a specific page image, hidden text and metadata referring to the source publication, the date of publication, page number, the reel number and sequence order.
- One OCR text file for each newspaper page image. The text conversion process will produce files that meet the specifications listed in Appendix B of the NDNP Technical Guidelines.

Each text file will contain:

- Uncorrected text
- Word-bounding boxes zoned for column recognition. Files will be free of article level segmentation.
- Bounding box coordinate data at the word level.
- UTF-8 characters
- No graphic elements
- The text created through OCR will be encoded using ALTO Version 2.0. If possible, the vendor will supply confidence level data at the page, line, character, and/or word level. Additionally, the vendor will seek to provide point size and font data at the character or word level.

Quality Control, Validation, Loading and Archiving

UF will remain in frequent communication with the vendor via telephone and email. After completion of microfilm scanning, file creation and metadata encoding. Vendor staff will validate the vendor created files. Upon completion of the validation process, the vendor will send validated

files on external hard drives to the Smathers Libraries where project staff will:

- Validate all deliverables (request vendor recreates deliverables that failed to validate); confirming vendor has accounted for discrepancies noted by project staff during their initial evaluation of reels
- Ensure vendor correctly used Issue Present and Page Present indicators
- Verify the four digital files associated with a newspaper page (TIFF, JP2, PDF and OCR text file) use the same file name and differ only by respective file extensions.
- Load all files online to the UF Digital Collections powered by the UF system
- Archive all files for long-term digital preservation

Appendix 1: UF Master Microfilm Inventory

Title	Dates on film, selected	OCLC	ISSN
Le Propagateur	1855	472564305	2428-1999
	1868		
	1870		
	1876		
	1881-1883		
	1885-1893		
Journal officiel de la Guadeloupe	1852, 1854-1855	1455443	1141-0701
	1858-1859		
	1860-1862		
	1863-1865		
	1866-1868		
	1869, 1871-1872		
	1873-1875		
	1876-1878		
	1879-1881		
	1885-1887		
	1888-1889		
	1890-1891		
	1892-1893		
	1899-1900		
	1906-1907		
	1908-1909		
	1910-1913		
1914-1917			
1918-1920			
1921-1929			