**LAMP Digitization Proposal**

*Text below in italics is directly from the LAMP Digitization Project Principles (*[*http://www.crl.edu/area-studies/lamp/news/proposal-guidelines*](http://www.crl.edu/area-studies/lamp/news/proposal-guidelines)*).*

*Standard information for all proposals from the University of Florida (UF) George A. Smathers Libraries is provided below, when applicable for all projects. This information is current as of July 2012.*

**I. Narrative**

**Title & Abstract**

*Diario de Pernambuco Project Phase II*

At the Trinidad LAMP meeting (June 16, 2012) UF proposed and received support to scan and digitize the *Diario de Pernambuco*. The University of Florida’s (UF) holdings include 276 reels of microfilm dating from 1825 through 1923. These reels are the only holdings outside of Rio’s Biblioteca Nacional, which was UF’s original source. UF is committed to hosting the digital newspaper content. Project cost estimates to digitize UF’s holdings of this important primary resource range from $52,000 to $100,000 depending on the number of pages on each reel. During the meeting in Trinidad, the membership voted to award UF $25,000 to cover the initial part of the project. This proposal respectfully requests an additional $25,000 to continue this work.

The following is a progress report as of May 6, 2013:

The George A. Smathers Libraries at the University of Florida, Digital Library Center (DLC) staff has processed 60 digitized reels returned from Creekside Digital. These 60 reels represent 10,028 issues of this newspaper dating from 1825 to parts of 1863. Access is free & open at:

<http://ufdc.ufl.edu/results/?t=diario%20de%20pernambuco>

DLC project staff worked with the vendor to precisely expend the grant award allocation.

**Content**

The *Diario de Pernambuco* is acknowledged as the oldest newspaper in circulation in Latin America (see: Larousse cultural; p. 263). Digitized newspapers during the proposed timeframe will offer researchers insights into early Brazilian commerce, social affairs, politics, family life, slavery, and such; published in the port of Recife. The *Diario* contains numerous announcements of maritime movements, crop production, legal affairs, and cultural matters.

The 19th century newspapers include reporting on the rise of Brazilian nationalism as the Empire gave way to the earliest expressions of the Brazilian republic. The 1910s and 1920s were years of economic and artistic change, with surging exports of sugar and coffee pushing revenues which supported rapid expansions of infrastructure, popular expression, and national politics.

**Copyright / Permissions**

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 (<http://dloc.com/AA00002865> and <http://ufdc.ufl.edu/permissions>). 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 to be provided based on the specific project needs.

**Conversion Procedure**

The UF Digital Library Center (DLC) is a one of the largest digitization and digital curation facilities in the southeast. The DLC utilizes many types of equipment and relies on industry standards for digitization that adheres to digital preservation standards.

The common workflow is shown in the image below.



All DLC imaging is completed in accordance with established professional standards. Imaging methods will depend on object characteristics, and follow principals and guidelines established in *Moving Theory into Practice: Digital Imaging for Libraries and Archives* by Anne R. Kenney and Oya Y. Rieger and Cornell University's *Digital Imaging Tutorial*. Imaging (i.e., scanning, text, metadata) is based on specifications previously established by UF and its partners (<http://digital.uflib.ufl.edu/technologies/documentation/imaging.htm>).

All objects are digitized to meet standard requirements for the item’s physical format. Images are captured as uncompressed TIFF files (ITU T.6) at 100% scale. All project imaging is calibrated regularly to maintain color fidelity and optimum image results.

Equipment for digitization includes:

* Super 8K-HS digital camera (for maps, architectural drawings and other large format materials)
* CopiBook (appropriate items, up to 15‖ x 23‖ sizes)
* Flatbed scanners (Microtek 9800 XL)
* Nikon Super CoolScan 5000 ED Film Scanner and Nikon SF-210 Auto Slide Feeder (slides, scanned individually or in batches)
* Details on all available equipment are here: <http://digital.uflib.ufl.edu/technologies/technologies.htm>

***What quality control will be used to ensure best practices are adhered to throughout the conversion process?***

UF’s DLC utilizes many types of equipment and relies on industry standards for digitization that adheres to digital preservation standards.

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

OCR text is uncorrected.

**Metadata**

Metadata processing is common for all materials.

Metadata: Metadata Encoding and Transmission Standard (METS; <http://www.loc.gov/standards/mets/>) metadata is created using the SobekCM tools and system, which are a full suite of production, digital collection (access), and repository (preservation) tools. The production workflow is integrated with the access system for consistency. As items are processed, the metadata is enhanced automatically and manually as objects move through the imaging/curation workflows. The SobekCM system assigns a unique Bibliographic Identifier (BibID) to each object processed, and that BibID is used to track the item (see UF Metadata Information, <http://ufdc.ufl.edu/sobekcm/metadata>). 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 SobekCM 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.

The SobekCM system stores all metadata in METS/MODS as well as automatically transforming and providing the metadata in MARCXML and Qualified Dublin Core, with all metadata accessible online. All materials are optimized for search engine access to ensure worldwide reach through Google and other search engines. SobekCM includes integrated support for OAI-PMH ([Open Archives Initiative or OAI)](http://www.openarchives.org/) to ensure all metadata is harvestable following OAI-PMH standards.

The SobekCM system specifications are optimized for data exchange for harvesting by other digital libraries such as the U.S. National Science Foundation’s National Science Digital Library, the U.S. Institute for Museum and Library Services’ National Leadership Grant collection, and OAIster at the University of Michigan.

**Added-Value Features**

***Describe any proposed products beyond digital image files. For instance:***

* ***Will text files be made searchable via the application of Optical Character Recognition software or double-keying?***
  + SobekCM provides 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?***
  + TEI and other schemas are applied on a project-specific base.
* ***Will numerical files be rendered in forms suitable for statistical manipulation?***
  + SobekCM supports standardized file formats, including data sets and numerical files.
* ***Will cartographic and related materials include geospatial referencing?***
  + Yes. SobekCM supports map-based searching and browsing for all materials with geographic metadata.

**Access**

***Describe how the users will access the data.***

***Delivery system:***

* + SobekCM, <http://ufdc.ufl.edu/sobekcm/>.
* ***In what format will the files be delivered?***
  + Imaged object files are delivered online in JPG, JPG2000, and JPG thumbnail images along with the OCR text files (TXT and PRO, for location of text on the image files) and with the metadata, displayed as a “citation” and also available and displayed in all metadata formats (METS/MODS, MARCXML, Qualified Dublin Core).
* ***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?***
  + SobekCM support for all collections and items includes:
    - Full text searchable
    - Browseable - with browse views by title and thumbnail, and by new items
    - 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, with automatic transformations for maximum interoperability
    - [Google-map based searching](http://ufdc.ufl.edu/aerials/map) or [map browsing](http://ufdc.ufl.edu/fdnl1/geography/)
  + Custom views for specific item-types:
    - [Full-screen page turner view](http://ufdc.ufl.edu/l/UF00028233/00001/pageturner)
    - 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.   
    SobekCM includes integrated support for OAI-PMH ([Open Archives Initiative or OAI)](http://www.openarchives.org/) to ensure all metadata is harvestable following OAI-PMH standards.

**Archiving**

***Describe terms for the preservation and ongoing maintenance of content.***

***What is your process for sustained preservation of the files?  
Will the data be archived at any location(s) other than CRL?***

The University of Florida George A. Smathers Libraries are committed to long-term digital preservation of all materials in the UF Digital Collections, including the IR@UF, and in UF-supported collaborative projects as with the Digital Library of the Caribbean (dLOC). 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 200,000 digital objects with over 20 million files (as of September 2011). 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.

In practice consistent for all digital projects and materials supported by the Libraries, redundant copies are maintained for all online and offline files. The digital archive is maintained by the Florida Center for Library Automation (FCLA). Completed by the FCLA in 2005, the Florida Digital Archive (FDA) (<http://fclaweb.fcla.edu/fda>) is available at no cost to Florida’s public university libraries. The software programmed to support the FDA is modeled on the widely accepted Open Archival Information System. It is a dark archive and no public access functions are provided. It supports the preservation functions of format normalization, mass format migration and migration on request.

As items are processed into the UF Digital Collections (UFDC) for public access, a command in the METS header directs a copy of the files to the Florida Digital Archive (FDA). The process of forwarding original files to the FDA is the key component in UF’s plan to store, maintain and protect electronic data for the long term. If items are not directed to load for public access, they do not load online and are instead loaded directly to the FDA.

***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. There are occasions when digitization for digital preservation is an absolute necessity because materials are disintegrating and cannot be preserved further in physical form. Most often, digitization for digital preservation is conducted alongside conservation of the physical materials where the materials, once conserved and if handled less frequently, will remain preserved in physical form. Because digitization for digital preservation and the ongoing work for digital curation are laborious and expensive processes, the physical objects selected for digitization are often from special and area studies collections where the physical materials are significant as artifacts and will continue to be preserved in that form.

**II Plan of Work**

***A detailed workplan should include an estimated schedule for the digitization project, broken down by the phases of the project (selection, permissions, preparation, conversion and quality control, metadata creation, delivery, preservation, etc). The workplan should also include information about the staffing needed to complete all aspects of the project.***

The UF DLC, as was done in Phase I, will work with Creekside Digital for vendor digitization. The various stages of metadata and quality control will follow. Files will be prepped for processing and loaded to UFDC. Additional details are provided in the budget estimates below. A dedicated webpage in the UFDC has been developed for supporting this project: <http://ufdc.ufl.edu/AA00011611>

**III. Budget**

***A detailed budget should include estimated costs for the digitization project, broken down by the phases of the project. The budget should include any project support requested of LAMP, as well as expected from sources other than LAMP.***

This will parallel last year’s Phase I budget.

|  |  |
| --- | --- |
| **Diario de Pernambuco Phase II estimate without Sloan subsidy 1875 to ~1893 - (83 Reels)** |  |
| **Expense Categories** | **Expense Detail** |
| (83) 1-up reels (650 frames each) 53950 frames/pages | 83 x 650 x $0.27=**$14,566.50** |
| Frame cost | Total frame cost = **$14,566.50** |
| Segmentation cost | Monthly segmentation (83 reels @ 3 monthly segments each) = 249 x $0.88 = **$219.12** |
| Combined frame and segment costs | Vendor services total = **$14,785.62** |
| OPS Labor: DLC preparation, UFDC ingest, Archiving to Florida Digital Archive | OPS labor $10/hr @ 11.94 hours per reel  (83 reels x 118.83/reel) = **$9,910.20** |
| Total vendor and labor costs | **$24,695.82** |
| Cost per reel = $297.54 |  |
| Shipping costs | **$300.00** |
|  | **Total cost = $24,995.82** |