Digital Border-Crossing: Practical Advice for Digitization Projects that Transcend Borders

Central American Broadsides
Kenneth Spencer Research Library
University of Kansas
http://lib.ku.edu/broadsides/

- Elspeth Healey, Special Collections Librarian
- Erik Radio, Metadata Librarian
- Betsaida Reyes, Librarian for Spanish, Portuguese, Latin American and Caribbean Studies
- Jocelyn Wehr, Digitization Services Coordinator
Digital Border-Crossing: Practical Advice for Digitization Projects that Transcend Borders

Session Outline

- Concept and criteria
- Collaboration
- Imaging
- Metadata
- Lessons learned: collaborating across borders

http://lib.ku.edu/broadsides/
William J. Griffith Collection of Guatemala and Central America

Acquired in 1982, the William J. Griffith Collection of Guatemala and Central America is the fruit of some forty years of collecting by the KU Professor Emeritus of Central American history. The collection consists of books, newspapers, broadsheets, printed ephemera and manuscripts, primarily from Guatemala and other Central American countries, with particular emphasis on the period from the late 1700s to the 1920s.

The collection is particularly strong for:

- printed ephemera (hundreds of broadsides and pamphlets)
- regional Guatemalan newspapers
- the history of education in Guatemala

Information regarding the related Griffith Manuscript collection is available in the "Finding Manuscripts by Country" portion of this guide.

SEARCH TIP:

- To browse titles in the Griffith collection, select the option "left-anchored call number" from the search settings in the KU's online catalog, enter "Griffith" and then hit "search." You may then sort the results by date, title, or author.

Image: Contestacion al Folleto Titulado: Manifiesto, Fundado en Derecho, que en 21 de Noviembre Ha Publicado el Licenciado Antonio Ortiz Umuela, Guatemala: Imprenta de la Universidad, 1840. Call Number: Griffith B384
Toward Digital Repatriation

“While some may assume on first glance that the digital object—as a surrogate—is meant to replace the physical object, no one, standard definition, nor agreed-upon terminology, characterizes the multiple practices of collecting institutions, individuals, or local community groups surrounding the return of cultural and historical materials to indigenous communities.”

Philosophies and Practicalities: Delimiting Your Project

Selection Criteria:
- Historical significance
- Scarcity
- Preservation
- Metadata availability
- Platform-friendly
- Enhanced functionality and adding value
- Room for Growth

http://lib.ku.edu/broadsides/
Selection Criteria: **Historical Significance**

- **National and International Politics**: Pronouncements by and about major political figures such as Guatemalan Presidents Mariano Rivera Paz (1804-1849), Rafael Carrera (1814-1865), Vicente Cerna (1815-1885), Justo Rufino Barrios (1835-1885), and Manuel Estrada Cabrera (1857-1924), as well as Honduran President Luis Bográn (1849-1895).
Selection Criteria: **Historical Significance**

- **Insight into Everyday Life:** mayoral politics, political clubs, theatrical productions, religious organizations
Selection Criteria: **Scarcity**

- Providing access to primary source materials not readily available elsewhere
- Associated responsibility of sharing these with their countries of origin and audiences around the globe


*KU’s copy is the only holding listed in WorldCat; Not in CCILSA or Metabase*
Selection Criteria: Preservation

- Ensuring the persistence of the informational content of broadsides that have suffered past damage (tears, folds, water, and insect damage)
- Reducing the necessity of handling fragile materials
Selection Criteria: **Metadata availability**

- Item-level metadata available
- Easily exported as MARC records
Selection Criteria: **Platform-friendly**

- Work with the technological constraints of your platform: broadsides can be captured as single images and therefore play well with the LUNA delivery system.
Selection Criteria: Enhanced functionality and adding value

- Good candidates for OCR, which though not presented as a reading text, could be used to provide additional discoverability via full text searching*

*Reality check to follow...
Selection Criteria: Room for Growth

Possibilities for expanding the digitization project:
- Incorporate yet-to-be cataloged broadsides
- Move beyond broadsides to pamphlets and books from the same collection
- Reach out to other institutions with strong holdings in the same subject area and invite them to participate in a shared portal
XI International Book Fair of Guatemala
FILGUA: Using the book fair as a springboard for other projects

- Taking advantage of government-sponsored trips to the book fair
- Exploring the country’s new publications and publishing industry
- Networking
“Returning” documents to their country of origin

Biblioteca Nacional de Guatemala
National Library of Guatemala
Visiting Guatemala’s National Library
Imaging: Preparing the broadsides

- Conservation Services department inventoried the condition of the broadsides prior to digitization:
  - Already housed in oversize folders
  - Some placed in mylar sleeves
  - Insect damage about 5%
  - Inactive mold about 5%
  - Majority of treatment for brittleness, tears, creases, and tape
  - Mended with Japanese paper and wheat starch paste
  - Forewarned by Conservation staff about potential hand irritation
Imaging: Preparing the broadsides

- What if you don’t have staff trained in conservation?
  - Could be an elective step if no active mold or infestation is present
  - Outsource work to NEDCC or Gerald Ford Conservation Center
  - Invest in archival folders and acid-free boxes
  - DIY humidification chamber
  - Use sheets of glass to flatten while imaging (depending on condition)

- Physical materials were identified with call number.
Imaging: Capturing the broadsides

• Equipment Used:
  o PhaseOne 645DF with P30+ digital back
  o 9’ camera stand
  o 4 Buhl softcubes

• What if you don’t have fancy equipment?
  o Use a flatbed scanner (like an Epson 10000XL)
  o Use a planetary scanner (like a Bookeye)
  o Camera mounted on table-top copy stand (shooting vertically)
  o Camera mounted on tripod (shooting horizontally)
Imaging: Capturing the broadsides

- File Naming
  - Need to establish consistent convention
  - Ensures that each file is unique
  - Detailed vs. numerical
  - ksrl_sc_griffith_h121_0001.tif (manual)
  - 00000001.tif (automated)

- Time Requirement
  - 1 student @ 12 hours/week X 8 weeks = 600 images
Imaging: Conducting Quality Assurance (QA)

• Essential step in the digitization workflow
• Every image vs. sampling
• Quality vs. quantity
• Professional workforce vs. student workforce
  o Expectations vary
• Problems we encountered:
  o Changing exposure/lighting
  o Inconsistent cropping
  o Camera shake
• Importance of documentation
Digitization Work Log

When

What

How

Who

Where

Priority (circle one): Standard  Rush  Needed By: 4/30/15

Title: Broadsides

Call Number: Griffith H101 - H124

Physical Description: oversize folders in 1 oversize box

Project: Central American Broadsides

Imaging Instructions: Adjust camera height to fill frame, crop w/slight border

Prepared By: J. Wehr

Imaged By: Tim Churchill  Date: April 21st 2015  Total Hours: 3 hours

Workstation: Phase 1  # of Images: 20 images

Path to Masters: Desktop Broadsides output.

DPI: 300  ISO: 100  Aperture: f8.  Shutter Speed: 1/30

Mode (circle one): Bitonal  Greyscale  RGB

QA’d By: J. Wehr  Date: 4/22/15  Total Hours: 1/8

QA Comments: Approved, OCR completed.

Date Loaded to Digmaster: 8/13/15 gw
Imaging: Conducting OCR

- Optical Character Recognition
- Adobe Acrobat Pro XI
- Other software options include Abbyy FineReader or built-in software
- OCR in Spanish
- Success rate of approximately 80%
- Sampled approximately five words per page
- Italicized, elaborate, or handwritten text make OCR difficult
- If you have staff or volunteer hours available, consider manually transcribing instead
- Apply for grants that allow you to hire a transcriptionist
Metadata: Processing

- MARC to MODS (LC stylesheet)
- Legacy data, granularity, controlled vocabularies
- Linked data
- Cataloger identity
- Metadata staging/freezing

```xml
<mods:subject authority="lcsh">
  <mods:name type="corporate">
    <mods:namePart>Catholic Church</mods:namePart>
  </mods:name>
  <mods:topic>Controversial literature</mods:topic>
</mods:subject>

<mods:subject authority="lcsh">
  <mods:topic>Church and state</mods:topic>
  <mods:topic>Catholic Church</mods:topic>
</mods:subject>
```
Metadata: System

- MARC to MODS to Luna (?)
- Semantic/Structural loss
- Elements in English
- OCR
- Defining users and needs

```xml
<fieldGroup type="Subject">
  <field type="Subject">González Vigil, Francisco de Paula, 1792-1875</field>
  <field type="Subject">Defensa de la autoridad de los gobiernos y de los obispos contra las pretensiones de la Curia Romana</field>
  <field type="Subject">Catholic Church</field>
  <field type="Subject">Controversial literature</field>
  <field type="Subject">Church and state</field>
  <field type="Subject">Catholic Church</field>
</fieldGroup>
```
Lessons Learned: Collaborating across Borders

- **Embrace exploration**: start with a pilot project to discover what works and doesn’t work
- **Select a project manager**: that person will help establish and communicate the workflow, timeline, and procedures
- **Take advantage of project management tools**: this will help facilitate communication and eliminate the danger of some project collaborators not knowing about changes that affect their work (e.g. reassigning of some call numbers)
- **Establish a shared consensus on project features early on**: if we had known that integrating OCR wouldn’t be possible, we would have enhanced the metadata (especially subject headings) to improve searchability
- **Be flexible**: Some things *will* change so be agile and ready to adjust
Questions?

http://lib.ku.edu/broadsides/

Elspeth Healey (ehealey@ku.edu)
Erik Radio (radio@ku.edu)
Betsaida Reyes (breyes@ku.edu)
Jocelyn Wehr (jocelyn@ku.edu)