

Rapid Digitization of Latin American Ephemera with Hydra

Project History and Goals

Princeton University Library began to collect and build an archive of Latin American ephemera and gray literature in the mid 1970s to document the activities of political and social organizations and movements, as well as the broader political, socioeconomic and cultural developments of the region. Access to the material was provided by slowly accumulating and organizing thematic sub-collections, creating finding aids, and microfilming selected curated sub-collections. Reproductions of the microfilm were commercially distributed and resulting royalties were used to fund new acquisitions. That model gradually become unsustainable during the past decade and microfilming was halted in 2008.

Hydra breathes new life into this project by providing us with a framework for creating an end-to-end application that will facilitate rapid digitization, cataloging, and access to this important collection. Since the system went into production in April of 2014, nearly 1500 items have been cataloged, with the throughput rate ultimately accelerating to over 300 items per month in August.

Princeton has several projects and workflows similar to "LAE", and we expect many of the components of this applications to reusable in future projects.

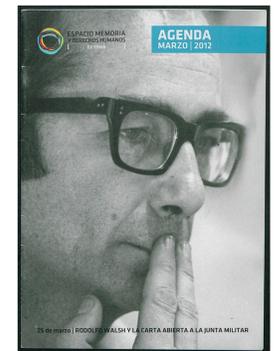
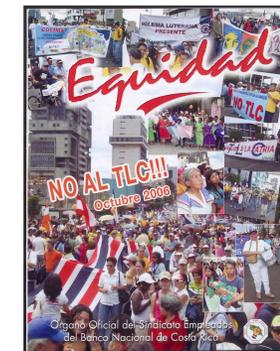
Rich Metadata Available as RDF/Linked Data

```
@prefix dc: <http://purl.org/dc/terms/> .
@prefix isolang: <http://id.loc.gov/vocabulary/iso639-2/> .
@prefix lcco: <http://id.loc.gov/vocabulary/countries/> .
@prefix lcga: <http://id.loc.gov/vocabulary/geographicAreas/> .
@prefix lcsh: <http://id.loc.gov/authorities/subjects/> .
@prefix marcrel: <"http://id.loc.gov/vocabulary/relators/"> .
@prefix pulst: <http://pul-store.princeton.edu/> .
@prefix pulterms: <http://princeton.edu/pulstore/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix tgm: <http://id.loc.gov/vocabulary/graphicMaterials/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
puls:00b84 dc:title "Memoria 2003."@es;
marcrel:mfp lcga:cl;
pulterms:heightInCM "21"^^xsd:int;
pulterms:isPartOfProject pulst:0005p;
pulterms:pageCount "15"^^xsd:int;
pulterms:widthInCM "15"^^xsd:int;
dc:coverage lcga:cl;
dc:created "2003"^^xsd:Year;
dc:format tgm:tgm007415;
dc:language isolang:spa;
dc:publisher "Asociación para la Cooperación en el Sur (ACSUR)-Las Segovias";
dc:rights "This digital reproduction is intended to support research, teaching, and private study. Users are responsible for determining any copyright questions."
@en;
dc:subject lcsh:sh85040810,
"Economic policy--Social aspects"@en,
"Economics--Political aspects"@en,
lcsh:sh85067385 .
```

```
puls:004kr dc:title "The Workforce of Puerto Rico: Fringe Benefits."@en;
marcrel:mfp lcco:pr;
pulterms:heightInCM "22"^^xsd:int;
pulterms:isPartOfProject pulst:0005p;
pulterms:pageCount "17"^^xsd:int;
```

[...]



Item Cataloging and Imaging Workflow

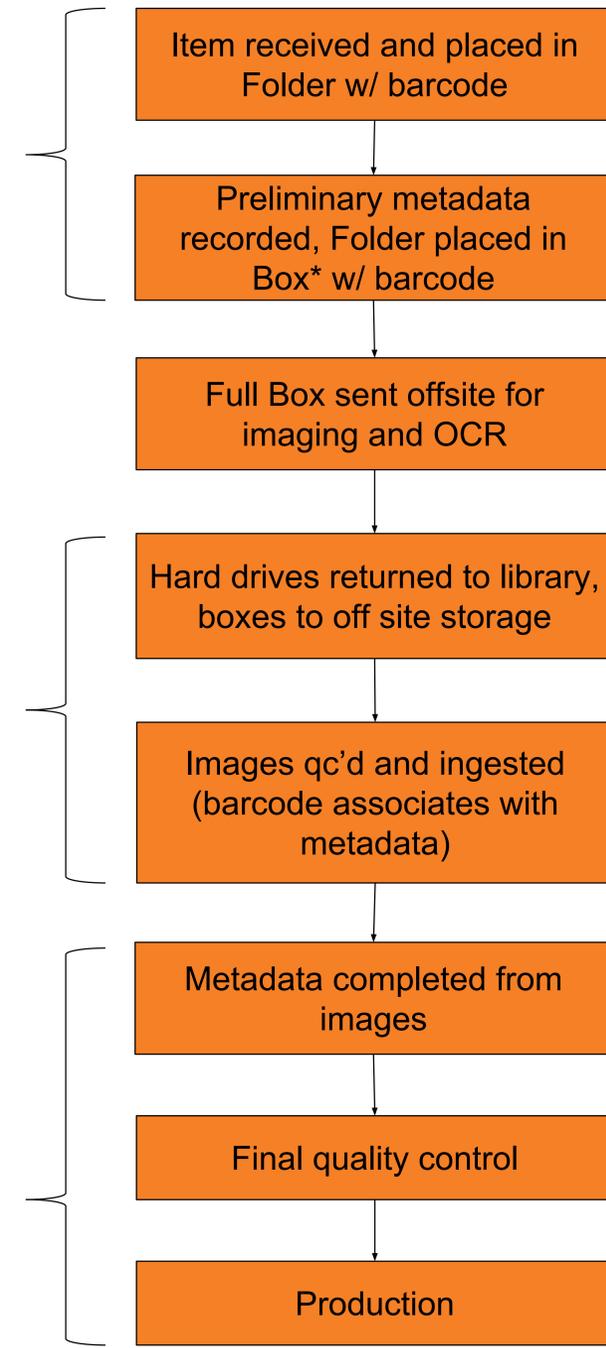
A Folder is instantiated in repo. Folder barcode, box barcode (via scanner), title, country, and genre MUST be included before it can be saved.

State is set to "Has Preliminary Metadata"

Returned drive is organized by box/folder barcodes. It also contains a JHOVE audit with checksums

Folders move through their remaining states:

- "Has Core Metadata"
- "Needs QC"
- "In Production"



* Boxes have a similar workflow. States are:

1. "New"
2. "Ready to Ship"
3. "Shipped"
4. "Received"
5. "All in Production"

Image QC includes:

- Checksums verified
- Color profiles confirmed
- Confirm (1) OCR for (1) file

Process returns a file of arguments which:

- Instantiating Page (image) and Folder objects
- Creating the appropriate associations
- This process also creates a JP2 that is copied to our image server (Loris).