Scanning for Scholarship

Digitization of State and Local Government Publications to Support New Frontiers of Original Research

Beyond the Numbers
October 7, 2016
St. Louis Federal Reserve Bank

Kris Kasianovitz
Government Information Librarian
krisk11@stanford.edu, @govinfogal
Data waiting to be unleashed!

State (and Local) Government Information

10/26/2016  @govinfogal
Stanford Library Digitization Services

The mission of Stanford University Libraries' digitization services is to further teaching, learning, and research at Stanford by using digitization to support the following library functions:

- Collection development (through digital acquisition, digitization as a condition for donation)
- Public services (patron requests, exhibits, items for class use)
- Technical services (preservation, digital surrogates for high use or fragile items)
- Research support (digitization in support of text mining)

Digitization is provided through four broad service areas:

- Project-based digitization – digitization of a library collection and large enough to require a project manager and significant coordination (Reid Dennis Collection of California lithographs)
- Patron and one-off internal requests – a single image or small set of items requested by SUL staff on behalf of a patron
- Systematic digitization – materials digitized en masse, typically for logistical or programmatic reasons (Glen McLaughlin Maps Collection — Maps of Malta)
- Consulting and outsourcing management – expert advice to other campus groups or in support of digitization work performed by vendors in support of Stanford projects (Bassi–Veratti Collection)

In addition to generating high quality digital surrogates of the original item, digitized items are typically accessioned into the Stanford Digital Repository (SDR), making them discoverable (findable) in SearchWorks (Stanford’s University Libraries’ online catalog). Inclusion in the SDR ensures that these materials are available to researchers and scholars (while upholding appropriate access restrictions), now and in the future through a secure, sustainable stewardship environment. Digitized items receive a PURL (a persistent URL), which ensures that these materials are available from a single URL over the long term, regardless of changes in file location or application technology.

http://library.stanford.edu/research/digitization-services/services
A Tale of Two Projects

State Banking Reports

- 3 Faculty and 1 Graduate Student from the Graduate School of Business
- Research: study of banks pre and post 1930’s Great Depression
- State Banking Commission Reports from 9 States: Arkansas, Georgia, Iowa, Kansas, Missouri, New Jersey, New York, Pennsylvania, and Virginia (complete runs from 1920-1940)
- 43 volumes*; 4,315 “images”, primarily all tables (*not available via Hathi, agency websites)
- Possibly copyrighted
- Table Re-keying outsourced

ICMA Municipal Yearbook

- Political Science Faculty
- Research: factors that impact city governance, differences in cities with council-manager vs. mayor-council; city finance and budget tables.
- 1 Title, 84 volumes, approximately 45,500 “images”
- Combination of text and tables
- Copyrighted by ICMA
- Table Re-keying by research assistants

10/26/2016 @govinfogal
Banking Reports: Publications are all over the place!
<table>
<thead>
<tr>
<th>state</th>
<th>year</th>
<th>Status</th>
<th>Hathitrust link</th>
<th>Printed sources</th>
<th>Wordcat link</th>
<th>NOTE</th>
<th>Name in original file</th>
<th>Updated Name by FA</th>
<th>Editon on Number of pages (from All Banks statistics)</th>
<th>source</th>
<th>source updated by FA</th>
<th>Number of pages (low number -&gt; pages missing from scanned document?)</th>
<th>Individually bank data in reports?</th>
<th>N state banks in report</th>
<th>other banks in reports</th>
<th>Officers listed</th>
<th>Directors listed</th>
<th>Ownership detail</th>
<th>Additional information that might be useful</th>
<th>Note</th>
<th>Hathitrust link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virginia</td>
<td>1920</td>
<td>Printed copies found</td>
<td><a href="http://verba.ill">http://verba.ill</a>. <a href="http://www">http://www</a>.</td>
<td></td>
<td>10/26/2016</td>
<td>@govinfogal</td>
<td>Anna 13</td>
<td>331 Hath Trust Covet not fit</td>
<td><a href="http://catalog.iit">http://catalog.iit</a></td>
<td>345 Hath Trust Covet not fit</td>
<td>37 1 ~ 334 0 0 0</td>
<td>Reduced number</td>
<td>Just capital on <a href="http://catalog.iit">http://catalog.iit</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>1921</td>
<td>Printed copies found</td>
<td><a href="http://verba.ill">http://verba.ill</a>. <a href="http://www">http://www</a>.</td>
<td></td>
<td>10/26/2016</td>
<td>@govinfogal</td>
<td>Anna 13</td>
<td>331 Hath Trust Covet not fit</td>
<td><a href="http://catalog.iit">http://catalog.iit</a></td>
<td>345 Hath Trust Covet not fit</td>
<td>37 1 ~ 334 0 0 0</td>
<td>Reduced number</td>
<td>Just capital on <a href="http://catalog.iit">http://catalog.iit</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Virginia | 1922 | Printed copies found | http://verba.ill. http://www. |     | 10/26/2016 | @govinfogal | Anna 13 | 331 Hath Trust Covet not fit | http://catalog.iit | 345 Hath Trust Covet not fit | 37 1 ~ 334 0 0 0 | Reduced number | Just capital on http://catalog.iit | | | | | | | | | |}

Banking Reports: Publications are all over the place!
Copyright & Permission

State Banking Reports

- Copyright Law Section 105 Covers Federal Government Information, not States!
- Post-1923 *could be* copyrighted
- We seek permission
- Work of Free State Gov Info http://stategov.freegovinfo.info/

Municipal Yearbook

- Clearly Copyrighted
- Negotiation process in-progress
My colleagues and I are working on a research project on “Banking during the Great Depression” that uses data from your agency’s Annual Reports from the years 1920 to 1940.

To aid in that research, we request your permission to digitize possibly copyrighted editions of those works, to archive the digital version of the works, and to make the digital works publicly accessible.

If you agree, please indicate your permission by responding to our email and including the information below:

The [insert state name] Bank Department hereby authorizes the Stanford University Library to digitize the [insert state] state publications listed below ("the Works"); to archive and preserve the Works in a digital archive, such as the Stanford Digital Repository; and to make the Works accessible to the public through its catalog, its website, or other publicly accessible tools, without restriction.

We further authorize Stanford University Library to share the digital versions of the Works with other digital archives or institutional repositories, including, but not limited to, HathiTrust.
2014 Municipal Year Book
The premier resource for local government information.

Regular Price: $156.00 | MEMBER PRICE: $120.00

To order, please call 770-280-4171; e-mail icma-orders@pbd.com; or fax 770-280-0092.

This special edition of the Municipal Year Book marks the incredible accomplishments of local government management and explores the opportunities that await the profession. In its pages, you will find a foreword by Executive Director Bob O’Neill on the legacy of local government professionalism, a 100-year retrospective of social policy issues, and a review of ICMA’s evolution as an international organization. Survey-based articles focus on the state of the local government profession today, the growing role of intermunicipal cooperation in alternative service delivery, CAO salaries and compensation, and salaries and expenditures for police and fire personnel.
Unlocking the “data”: .tiff → OCR’d .pdf → .csv/.xls

- ABBYY FineReader for OCR
- Tabula?
- IBM Bluemix?

Additional steps:
- Manual re-key, with audit
- Manual re-key with arbitration
- Hire freelancers
- Hire undergrad/grad student assistants

https://purl.stanford.edu/qk950cx7608
## Stanford Digital Repository – Online Deposit

### Stanford Research Data

89 Items in This Collection

<table>
<thead>
<tr>
<th>Title</th>
<th>Depositor</th>
<th>Status</th>
<th># of Files</th>
<th>Date Added</th>
</tr>
</thead>
<tbody>
<tr>
<td>state government workers’ perceptions of their work</td>
<td>benditt</td>
<td>published</td>
<td>7</td>
<td>2016-09-25 4:12 pm</td>
</tr>
<tr>
<td>Energy–based Constitutive Modeling of Local Material Properties of Canine Aortas</td>
<td>klaksari</td>
<td>published</td>
<td>4</td>
<td>2016-08-23 9:03 pm</td>
</tr>
<tr>
<td>Rehydration Calculator Overview</td>
<td>rball</td>
<td>published</td>
<td>4</td>
<td>2016-08-18 9:31 pm</td>
</tr>
<tr>
<td>Data for submission of “How pigeons couple elbow and wrist motion to morph their wings”</td>
<td>astowers</td>
<td>published</td>
<td>1</td>
<td>2016-08-17 1:12 pm</td>
</tr>
<tr>
<td>Syncytia formation in melanoma cells infected with varicella zoster virus and the hyperfusogenic mutant gB[Y881F]</td>
<td>sloriver</td>
<td>published</td>
<td>5</td>
<td>2016-08-05 1:58 pm</td>
</tr>
<tr>
<td>Poverty mapping replication code</td>
<td>mburke</td>
<td>published</td>
<td>1</td>
<td>2016-08-04 9:34 am</td>
</tr>
<tr>
<td>PASCALRAW: Raw Image Database for Object Detection</td>
<td></td>
<td></td>
<td>0</td>
<td>2016-07-19 11:22 am</td>
</tr>
</tbody>
</table>
Thank You!
Questions? Comments...

Kris Kasianovitz, krisk11@stanford.edu
@govinfogal
Demo:
Scanning materials using the Atiz book scanner
Stanford Library DLSS “Robot Scanner”