Susan Bokern
Vice President, ProQuest
Government & Historical Collections

GOVERNMENT CONTENT: PRESERVED AND REIMAGINED
APRIL 25, 2014
Different Roles we all play
- Library of Congress
- NARA
- Government Agency Libraries
- Government Printing Office
- Intergovernmental Organizations
- GOOGLE
- Depository Libraries
- Academic Libraries
- National Libraries of countries around the world
- Aggregators
- Non Profits

Aggregators: ProQuest
- We have to create value or we don’t have a role
- We have to constantly keep updating our functionality and technology capabilities to meet emerging researcher needs and expectations
Role of government is to create and disseminate government information

Ironic: while U.S. Government announces ‘open data’ initiatives, it also ceases creating content

- Agencies not producing reports any longer
- When the government takes down content/data from its web sites = it’s gone
- U.S. Federal Government Sequestration = will this happen again?

Some of the Statistical Reports Discontinued in 2013:

- **This Month in Futures Markets** Source: Commodity Futures Trading Commission
- **Forest Health Monitoring Reports** Source: Forest Service
- **U.S. Contributions to International Organizations: Report to the Congress** Source: Department of State
- **Financial News for Major Energy Companies** Source: Energy Information Administration (Department of Energy)
- **Long-Term Care Ombudsman Report [yr.] with Comparisons of National Data for [yr.]** Source: Department of Health and Human Services
- **Telephone Subscribership in the U.S.** (Approx. 55 p.) Source: Federal Communications Commission
- **Computers and Peripheral Equipment. Current Industrial Reports** Source: Bureau of Census
Each agency does its own thing

Cuts to Census Bureau, BLS, USDA, BEA

BLS discontinues:
• Measuring Green Jobs
• Mass Layoff Statistics
• International Labor Comparisons

USDA suspends:
• Agricultural Productivity in the U.S.
• County-Level Data Sets ("County Look-Up Tables")
• Fertilizer Imports/Exports Organic Handlers: Procurement and Contracting
• Commodity Consumption by Population Characteristics

NTIS ceases:
• World News Connection

USDA discontinues:
• Reports and datasets on dairy, cattle, vegetables, and 7 other commodities

Census Bureau delays:
• Planning for the 2020 decennial census
• Processing and release of 2012 Economic Census

BEA curtails:
• Local Area Personal Income program
• Analysis of Foreign Direct Investment impacts
The Graveyard of Statistical Collections
The Census Reform Act

H. R. 1638

To repeal the authority to conduct certain censuses, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

APRIL 18, 2013

Mr. DUNCAN of South Carolina (for himself, Mr. CHAFFETZ, Mr. HARRIS, Mr. JONES, Mr. PEARCE, Mr. RUBLE, and Mr. SOUTHERLAND) introduced the following bill, which was referred to the Committee on Oversight and Government Reform, and in addition to the Committees on Agriculture and Appropriations, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned.

A BILL

To repeal the authority to conduct certain censuses, and for other purposes.

1 Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

2 SECTION 1. SHORT TITLE.

3 This Act may be cited as the “Census Reform Act of 2013”.

4

5
Data Collection is Costly

- The 1960 census cost $520 million (in 1990 dollars)
- The 1990 census cost $2.6 billion.
- The 2010 census cost $7.5 billion (in 1990 dollars)

* A 1300% increase in 50 years!
Federal Staffing Levels Are Flat

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>15,841</td>
<td>17,212</td>
<td>18,745</td>
<td>19,540</td>
<td>20,876</td>
<td>21,725</td>
<td>22,116</td>
<td>22,462</td>
<td>22,668</td>
<td>22,607</td>
</tr>
<tr>
<td>Federal (civilians)</td>
<td>3,023</td>
<td>3,191</td>
<td>3,507</td>
<td>4,009</td>
<td>4,899</td>
<td>3,690</td>
<td>2,963</td>
<td>2,963</td>
<td>2,824</td>
<td>3,068</td>
</tr>
<tr>
<td>State and local</td>
<td>12,818</td>
<td>14,082</td>
<td>15,238</td>
<td>15,531</td>
<td>16,977</td>
<td>18,035</td>
<td>19,153</td>
<td>19,146</td>
<td>19,068</td>
<td>19,599</td>
</tr>
<tr>
<td>Percent of total</td>
<td>81</td>
<td>82</td>
<td>84</td>
<td>86</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>88</td>
<td>88</td>
<td>87</td>
</tr>
<tr>
<td><strong>Local</strong></td>
<td>5,022</td>
<td>5,263</td>
<td>5,828</td>
<td>6,143</td>
<td>6,622</td>
<td>7,145</td>
<td>7,148</td>
<td>7,267</td>
<td>7,148</td>
<td>7,647</td>
</tr>
<tr>
<td>Counties</td>
<td>1,824</td>
<td>1,963</td>
<td>2,253</td>
<td>2,425</td>
<td>2,675</td>
<td>2,675</td>
<td>2,675</td>
<td>2,675</td>
<td>2,675</td>
<td>2,675</td>
</tr>
<tr>
<td>School districts</td>
<td>4,194</td>
<td>4,627</td>
<td>5,134</td>
<td>5,675</td>
<td>5,675</td>
<td>5,675</td>
<td>5,675</td>
<td>5,675</td>
<td>5,675</td>
<td>5,675</td>
</tr>
<tr>
<td>Special districts</td>
<td>356</td>
<td>393</td>
<td>424</td>
<td>455</td>
<td>455</td>
<td>455</td>
<td>455</td>
<td>455</td>
<td>455</td>
<td>455</td>
</tr>
<tr>
<td>PAYROLLS (mill. dol.)</td>
<td>478</td>
<td>529</td>
<td>627</td>
<td>691</td>
<td>691</td>
<td>691</td>
<td>691</td>
<td>691</td>
<td>691</td>
<td>691</td>
</tr>
</tbody>
</table>

**Table 469. Governmental Employment and Payrolls: 1982 to 2010**

[Employees in thousands (15,841 represents $15,841,000), payroll in millions of dollars ($23,173 represents $23,173,000,000). Data are for the month of October through 1992. Beginning with the 1997 survey, data are for the month of March. Covers both full-time and part-time employees. Local government data are estimates subject to sampling variation; see Appendix III and sources.]

NA Not available. Includes employees outside the United States.

GOVERNMENT CONTENT: PRESERVED & REIMAGINED

PROQUEST’S ROLE: GOVERNMENT CONTENT IN CONTEXT OF RESEARCHER NEEDS

- **Our focus:** is on researchers and the research workflow processes – to help them find, gather, store, save, create, share, and output their research

- Additional value comes from creating the discovery, metadata, search interface, tools – to meet researcher needs

Finding content, making it useful to create better research outcomes
GOVERNMENT CONTENT: PRESERVED & REIMAGINED

PROQUEST’S ROLE

1. Process begins with market research to uncover research needs
   • Surveys, advisory boards, focus groups, discussions with librarians and faculty

2. Create an acquisitions strategy to develop a collection, preserve the content or data,
   and keep the technology up-to-date for users to access it over time

3. Have to know where to obtain the content – which source locations
   • i.e., Executive Branch Documents (1789-1932) – 100 source libraries

4. Establish Criteria…varies
   • comprehensive to include “all” content:
     • i.e., Congressional, 1789-current, Parliamentary House of Commons Papers (1715-current)
   • selective, yet includes complete series of government publications
     • i.e., Statistical - 10,000 titles, 1M Tables
   • all relevant data from government statistical and data offices worldwide
     • i.e., International Datasets
   • comprehensive index that links to government content
     • i.e., Government Publications Index, GPI; NTIS index
100+ Professional Staff Expertise in government content/data

Value is the expertise of the staff: many with 35 years of service in...
- content acquisitions - worldwide
- authoritative bibliographic records
- editorial subject knowledge in abstracting and indexing
- legislative history research = professionally researched histories
- statistical/data expertise, why we could continue to create the
  - Statistical Abstract of the U.S.
GOVERNMENT CONTENT: PRESERVED & REIMAGINED

WHAT HAS CHANGED? OUR USERS!

• It’s a Digital World – Discovery & Use
  • Students, Faculty expect and want to find and use information at their fingertips wherever they are on whatever device they are using – PC, Mobile, Phone

• Students/Young Faculty – “all digital”
  • What’s microform? What’s a microform reader? You want me to do what?

• A&I moving to Full Text
  • Students don’t want to hunt for the full text separately, they want direct links to the full document

• Statistics turning into Data
  • Researchers want to manipulate and turn the data into original research output

• It’s no longer just about the document!
  • It’s about turning the document inside out to obtain new insights

• User expectations change as technology changes
  • New capabilities via Technology: Text Mining/Data Mining
IT’S NOT JUST ABOUT THE DOCUMENT ANYMORE!

It’s about…

• enabling researchers to access and work with the materials according to the technology available at the most current time: now and into the future

• providing tools to researchers to improve their workflow processes in the content gathering, storing and sharing stages

• simplifying researcher workflows = saving them time

• helping researchers make connections they couldn’t see before
  • = VISUALIZATION
  • = ANALYTICS
  • = TEXT MINING, DATA MINING

• offering cloud-based services with new business models
GOVERNMENT CONTENT: PRESERVED & REIMAGINED

NEW ROLES FOR PROQUEST: ADDING VALUE TO CONGRESSIONAL CONTENT
GOVERNMENT CONTENT: PRESERVED & REIMAGINED

NEW ROLES FOR PROQUEST: SOLVING COMMON STATISTICAL AND DATA PROBLEMS

I need to know how many or how much.

I need multi-year trends, deep time series.

I want to compare two countries, states, companies, industries, etc.

I want to manipulate the data.
Statistics from governments around the world.

Statistics from the U.S. March 2011, the Census Bureau announced that the 2012 edition would be the last one produced at government expense.
NEW ROLES FOR PROQUEST: MAKING INTERNATIONAL GOVERNMENT DATA DISCOVERABLE, EASY-TO-FIND & USE

Data from governments around the world. Saves time in the researcher workflow process in finding the data.
GOVERNMENT CONTENT: PRESERVED & REIMAGINED

NEW ROLES FOR PROQUEST: MAKING INTERNATIONAL GOVERNMENT DATA USABLE

Data from governments around the world.

Cost to start a business (% of income per capita)

Units: % of income per capita

<table>
<thead>
<tr>
<th>Year</th>
<th>Brazil</th>
<th>Canada</th>
<th>China</th>
<th>Russian Federation</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>18.1</td>
<td>0.9</td>
<td>13.6</td>
<td>8.2</td>
<td>8.8</td>
</tr>
<tr>
<td>2006</td>
<td>9.9</td>
<td>0.9</td>
<td>9.3</td>
<td>4.6</td>
<td>8.8</td>
</tr>
<tr>
<td>2007</td>
<td>10.4</td>
<td>0.9</td>
<td>8.4</td>
<td>3.7</td>
<td>8.7</td>
</tr>
<tr>
<td>2008</td>
<td>8.2</td>
<td>0.5</td>
<td>8.4</td>
<td>2.2</td>
<td>8.7</td>
</tr>
<tr>
<td>2009</td>
<td>6.9</td>
<td>0.4</td>
<td>4.9</td>
<td>1.8</td>
<td>8.7</td>
</tr>
<tr>
<td>2010</td>
<td>7.3</td>
<td>0.4</td>
<td>4.5</td>
<td>2.8</td>
<td>1.4</td>
</tr>
<tr>
<td>2011</td>
<td>5.4</td>
<td>0.4</td>
<td>3.5</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Source: World Bank (citing World Bank, Doing Business Project (http://www.doingbusiness.org/)

Providing a variety of display, analytical, and export options
Promoting the discovery of high-value content on hard-to-navigate government sites.

Examples:

- India parental preference for male vs female children, by sex and state, 2005-2006
- India Wages in Cotton Mills [monthly wage for lowest-paid operative, selected cities, from 2010]
- India women's mean age at effective marriage in major states, [by urban-rural residence, 2005-2011]
Again, the irony: while it announces ‘open data’ initiatives, it ceases creating content

- Should they preserve all their content? Where?
  - Preservation is a new concept to them.

- Should they be creators of published reports with analysis?

- Should they create robust web sites?

- Or, should they just provide the raw data?

- Where and how should they store researcher data sent back to governments in ‘open access models’?

- How is that content made available for additional research?
Some challenges:

• Terms of obtaining content from governments worldwide is evolving –
  • some agencies require licenses and some international ones seek royalties

• Aggregator costs are increasing significantly - acquiring, digitizing, adding value

• Government entities need to adopt practices that encourage aggregators to obtain a reasonable ROI

• For print/microform content, they need to reconsider embargo periods before digitized content is put on their web sites for free, or aggregators will not be able to afford to digitize it.

• For ‘born digital’ content, they need to make the content/data more accessible.

• If we can’t partner effectively and efficiently, the outcome:
  • less digitization, making fewer government documents available for research and teaching.

• Open to new business models.
NEW ROLES AND RESPONSIBILITIES FOR LIBRARIANS

- Faculty are approaching aggregators directly to obtain government and other content and metadata for research projects
  - Example: Matt Connelly/Bob Scott – Columbia U

- A growing disconnect between faculty needs/research and their perception of what the library provides

- Is there a role for librarians to?
  - Store government content/data that researchers gather and output as part of their research workflow process? Share it with aggregators?

- Librarians have told us = new role as facilitator
  - helping researchers, faculty find content for projects
  - explaining and managing copyrights
  - licensing materials = some government content around the world needs licenses
  - becoming more technically proficient to work with IT departments and help build capacity/sustainability year-over-year

- Open to new business models
WHAT I THINK ABOUT?

• Government Open Access / Open Data

• Perceived Free Content
  – i.e., Google ($400 Billion Market Cap = someone is paying)

• Curation = who is digitizing what content?

• Preservation = in what forms do we keep content? For the future?

• Migration of Print/Microform to Digital
  • Costs are enormous to change content into various digital formats

• New Technologies: new start ups, non-profits

• Libraries becoming publishers & mini-aggregators

• Disappearing Government Documents  librarian positions = folding into ‘Reference’

• Libraries & space considerations – trade offs

• Libraries & ROI’s to justify spend
New roles for us all!

Over time, the role of librarians in government aggregation and preservation will become clearer, as libraries

- determine which activities can best be done in-house
- determine which activities can be carried out most effectively and efficiently by aggregators

Let’s build that bridge together…to solve the challenges in supporting research.
• Thank You!

• Questions and comments

  • susan.bokern@proquest.com
  • 301-968-3820
Off the Beaten Path: Govdoc Frontiers

Leviathan: Libraries and Government Information in the Age of Big Data
Chicago • April 24-25, 2014

Robert E. Lee
Director of Strategic Partnerships and Online Publishing
East View Information Services