Bringing Data Home: Library and Vendor Perspectives on Purchasing Locally Hosted Data

Anita K. Foster, Electronic Resources Officer, the Ohio State University Libraries
Kris Hodgins, Manager, Gallup Analytics
Jeremy Groen, Academic Market Manager, Infogroup
Gene R. Springs, Collections Strategist, the Ohio State University Libraries

Charleston Conference, November 8, 2017
Focus on data at the University rapidly increasing since 2012

OSUL launches Research Commons (RC) services (2014) and its facility (2016)

Growing awareness of data as an information need across disciplines
Data Purchases

- Received requests for various datasets from different constituents
- Developed selection criteria during exploration of acquisition
- Discovered various models for data access
- Will these data potentially benefit the most researchers on our campus?
Hosting Considerations

- Asked for approximate size of the data files (first delivery and subsequent updates)
- Host (RC) agreed to provide server space for the data
- Delivery schedule included in license
Licensing Challenges

- Auditing concerns
- Export Control issue with one data purchase
- Time to process licenses
- Agreement/compromise on capabilities when relationship ends
About Us
Data Provided and Timeline

• Historical Business Data (1997 – 2016)
  • 268 million business records
  • Delivered in CSV & SAS format

• Historical Residential Data (2006 – 2016)
  • 1.55 billion residential records
  • Delivered in CSV & SAS format

• License Agreement process began in October 2016 and finalized end of February 2017

• Is this a typical timeline?
CHALLENGES

• Explaining why the license agreement reads the way it does
• Understanding university turn times for license agreements
• Wrangling our Legal team for manageable response times
• Managing/Hosting files of this size (Can you host this data in house?)
• Understanding the data provided
Licensing Gallup Data

Kris Hodgins - Manager, Gallup Analytics
Gallup Data Offerings

Three Major Polling Databases

1. World Poll
   - Nationally representative poll of over 140 countries conducted every year
   - Over 1.8 million interviews

2. US Tracker Poll
   - Nationally representative poll of the US conducted daily
   - Over 3.5 million interviews

3. Gallup Poll Social Series (GPSS)
   - Monthly poll focused on a rotating series of topics
   - 1,000 interviews conducted each month

Two Types of Data Access

1. Gallup Analytics Portal
   - Web-based tool that provides access to aggregate level data
   - Analyze data using standard demographic and geographic identifiers

2. Respondent Level Data
   - Files provided with all responses from every interview
   - Analyze data using a statistical package like SPSS or STATA
Gallup University Subscription Options

Single User Gallup Analytics Portal
- One concurrent user can access the portal
- IP Authentication

Campus Wide Portal Subscription
- Unlimited concurrent users can access the portal
- IP Authentication
- Includes 2 user licenses for the World Poll and US Tracker Poll respondent level datasets

Campus Wide Portal Enhanced Subscription
- Unlimited concurrent users can access the portal
- IP Authentication
- Includes unlimited user licenses for the World Poll and US Tracker Poll respondent level datasets
Thank You

Kris Hodgins
Manager
Gallup Analytics
Kris_Hodgins@Gallup.com
Follow us @GallupAnalytics
Copyright Standards

This document contains proprietary research, copyrighted materials and literary property of Gallup, Inc. It is for the guidance of your organization only and is not to be copied, quoted, published or divulged to others outside your organization. All of Gallup, Inc.’s content is protected by copyright. Neither the client nor the participants shall copy, modify, resell, reuse or distribute the program materials beyond the scope of what is agreed upon in writing by Gallup, Inc. Any violation of this Agreement shall be considered a breach of contract and misuse of Gallup, Inc.’s intellectual property.

This document is of great value to Gallup, Inc. Accordingly, international and domestic laws and penalties guaranteeing patent, copyright, trademark and trade secret protection safeguard the ideas, concepts and recommendations related within this document.

No changes may be made to this document without the express written permission of Gallup, Inc.

Gallup® is a trademark of Gallup, Inc. All rights reserved. All other trademarks and copyrights are property of their respective owners.
Post Purchase Activities

- Data Set File Delivery
  - Arrived within licensed and agreed upon timelines
  - Library IT staff downloaded and stored files while access methods were developed
Post Purchase Activities

- Working with Research Commons
  - License interpretation
  - Modes of Access
  - Researcher Referrals
  - Forwarding notifications of availability of new files
Post Purchase Activities

- **Complications**
  - Research Commons staff agreed to host data files on their server and manage access for researchers but it’s taking longer than expected.
  - Infrastructure of their computer labs and servers made access more complicated than expected.
  - Service Point expectations
    - Problem solving unique research needs vs. general access to the data file.
    - We’re all learning as we go.
      - Figuring out how best to serve up data by working directly with researchers interested in the data.
Takeaways

- Invite the Licensing person to vendor demos to provide a deeper base of understanding
- Involve IT in the selection process as soon as we know the data requires local hosting
- Include Head, Research Services in feasibility analysis for data hosted in the Research Commons
- Develop clear internal documentation for roles and responsibilities of hosting organization
Hosting Feasibility

1. Decision to Purchase Made
2. Order Placed
3. License Negotiated
4. Invoice Received
5. Resource Activated
6. Resource Cataloged
7. Listed on Database
8. Usage Statistics and Assessment
9. Decision to Purchase Made
10. Order Placed
11. License Negotiated
12. Invoice Received
13. Resource Activated
14. Resource Cataloged
15. Listed on Database
16. Usage Statistics and Assessment

Steps:
- Collection Manager
- Collection Strategist
- Electronic Resources Officer
- Quote and License from Vendor

Process:
- Hosting Feasibility

Flowchart:
- Hosting Feasibility
- Decision to Purchase Made
- Order Placed
- License Negotiated
- Invoice Received
- Resource Activated
- Resource Cataloged
- Listed on Database
- Usage Statistics and Assessment
Hosting Feasibility

Collection Manager → Collection Strategist → Electronic Resources Officer → Quote and License from Vendor

Local Hosting → RC & IT Staff → Feasibility of Hosting → Collections Strategy

Decision to Purchase Made

Order Confirmed → License Negotiated

Invoice Received → Resource Cataloged
Resource Activated → Listed on Database

Usage Statistics and Assessment
Conclusion

- Research Data Set purchasing will likely continue
  - Integrated into Collections Strategy processes
- Development of an assessment process
  - Who, what, when and how to count usage
  - Decide what “value” means for this type of acquisition
- Continue to refine best practices for data acquisition
Questions?

- Anita Foster – foster.1037@osu.edu
- Gene Springs – springs.14@osu.edu
- Jeremy Groen - Jeremy.Groen@infogroup.com
- Kris Hodgins - Kris_Hodgins@gallup.com

OSU Libraries: https://library.osu.edu/about/jobs/faculty
Cohort: Head, Bibliographic Initiatives, Metadata Initiatives, Acquisitions Librarian & Scholarly Sharing Strategist