Cat Among the Data: Encouraging Researchers to Deposit Data

Jen Doty and Rob O’Reilly
Emory University

Presentation at IASSIST & CARTO 2018
Background

• ICPSR Official and Designated Representatives for Emory University
• “The Sleuthful Data Steward” presentation from the 2015 ICPSR OR meeting: https://youtu.be/McBGP8zZ24c
• LEADS data provided when the ICPSR visited Emory in 2015 to run day-long data curation workshops
Our “LEADS” list

Sources*:

• NSF and NIH awards databases
• ICPSR shared data on Emory awards of interest from LEADS database
• ICPSR’s research: “Evaluating Impact of the 2009 American Recovery and Reinvestment Act on Social Science”
• ICPSR collection development policies

*links on final slide

Compiling our “LEADS” list

Ran queries on sources:

NIH

• initial query for awards with start dates >= 01/01/2006 produced over 4700 awards
• 6636 awards for a total amount of $2,591,000,957 when query was run to include 2005 and 2017

NSF

• 235 awards with start dates >= 01/01/2005 (initial downloads on 02/21/2016)
• 284 awards for a total amount of $116,220,623 when we ran query again to include 2017
Narrowing list

• limited to NSF awards for 2005 onward from the Directorate for Social, Behavioral & Economic Sciences (SBE): 70 awards
• focused on SBE for various reasons
  • Emory SBE recipients have closer ties to our library
  • Social & Behavioral Sciences orientation tends to align with ICPSR collection policies

Coding list

• PI department affiliations
• Pre- and post-2011 NSF mandate for DMPs
• ICPSR collections policies - priorities were highlighted
• Awards appeared in LEADS data shared by ICPSR
• Comments to further narrow list to best leads

Already limited on:
• Date ranges
• NSF directorate
<table>
<thead>
<tr>
<th>Award #</th>
<th>Title</th>
<th>Program</th>
<th>Start Date</th>
<th>Status</th>
<th>AB</th>
<th>AC</th>
<th>AD</th>
<th>AE</th>
<th>AF</th>
<th>AG</th>
<th>AH</th>
<th>AI</th>
<th>AJ</th>
<th>AK</th>
<th>AL</th>
</tr>
</thead>
<tbody>
<tr>
<td>055037</td>
<td>Doctoral Dissertation Improved</td>
<td>BCS</td>
<td>07/01/2006</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>0550387</td>
<td>Attitudes about Gender and the</td>
<td>LAW AND</td>
<td>07/15/2006</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Would like to talk</td>
<td>Yes</td>
</tr>
</tbody>
</table>
| 0655381  | OxoMetal Complexes and Red Cap
*Physical*                                                  | PHYSICAL | 03/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | File this one away. | File Away |
| 0029055  | Materials World Network: Fund                                          | DMR     | 06/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0050358  | Doctoral Dissertation Improved                                             | BCS     | 09/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | File this one away. | File Away |
| 0015409  | Rig: Co-orthogonal ABC Traf  
*Medical*                                      | MCB     | 09/15/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0616097  | Refined Mine Inhibitory Init
*Institute*                                               | IOS     | 08/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0616383  | Simulation and theory of mole  
*Chemical*                   | CHE      | 07/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | Beth was working    | Yes |
| 0618368  | Collaborative Research: Reproduc  
*Science*                        | POLITICAL | 06/15/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | Interesting        | No |
| 0622393  | Doctoral Dissertation Improved                                             | BCS     | 08/15/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR's cup of     | No |
| 0623951  | Integrating Models of Trust, GIIS
*Digital*                                      | DIGITAL  | 08/24/2005 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0625237  | Collaborative Research: Cyber
*Chemical*                   | CHE      | 09/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0626191  | Collaborative Research: Drug  
*Chemical*                        | PARTICUL | 09/01/2006 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0648776  | Doctoral Dissertation Improved                                             | BCS     | 07/15/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0648757  | Doctoral Dissertation Improved                                             | BCS     | 03/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0650437  | Doctoral Dissertation Improved                                             | BCS     | 04/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0653382  | Study of homogeneous space                                                | DMS     | 07/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0653502  | Semisimple groups and data                                                | DMS     | 07/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0653681  | Conference: Linear algebric                                              | DMS     | 04/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0703899  | NSF Inorganic Chemistry Work                                              | CHE      | 04/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0706985  | Quantile Regression With DMS                                              | STATISTIC | 07/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0711371  | Collaborative Research: Human
*Chemical*                        | BCS     | 08/15/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0717608  | Poverty, Social Change, and BCS                                           | BCS     | 09/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0718010  | Renewal: Social Mobility in the Americas                                 | BCS     | 07/15/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0719945  | Doctoral Dissertation Research Guide                                      | SOCIOLO | 09/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0720781  | CRS - AES Integrative Approaches                                         | CNS     | 08/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0723805  | Neuroendocrine Basis of Behav
*Social*                           | MODULATE | 09/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0724759  | CMG Collaborative Research: BCS                                           | DMS     | 09/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
| 0725269  | Collaborative Research: Integr
*Chemical*                        | HUMAN OR | 09/01/2007 |        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | ICPSR would be     | No |
And the winners are...
Subject librarian collaboration

• Involving subject librarians is an outreach “best practice”
  • Internal goal to further collaborate with our Scholarly Communications Office

• At Emory, we chose librarians who
  • participated in our RDM Task Force
  • are liaisons to departments with PIs on select awards

Case studies

Case Study #1

• Sociology professor, Irene Browne

• Background: deposited data with ICPSR in past; invites data librarian to course sessions; direct consultations for her research

• Brief award summary: qualitative data (interviews focusing on immigrant experiences with discrimination and mobility – relates to an ICPSR collection priority)

Case Study #1

• Sociology professor, Irene Browne
• Considerations and concerns: confidentiality, de-identification of data
• Outcome: existing data – not comfortable sharing for confidentiality reasons; *BUT* she was working on a new NSF proposal, so we connected her with ICPSR for discussion and letter of support to place any resulting data from that award in the ICPSR data archive
Case Study #2

• Political Science professor, Micheal Giles
• Background: award flagged for interest by ICPSR; invites data librarian to course sessions; direct consultations for his students’ research
• Brief award summary: coding En Banc court hearings for cases heard by Appeals Courts in their entirety

Case Study #2

• Political Science professor, Micheal Giles

• Considerations and concerns: approaching retirement, still cleaning data and preparing codebook, supplementing data, and planning future publications

• Outcome: didn’t decline, but preoccupied with all the above concerning the data
Case Study #3

• Anthropology professor, Craig Hadley
• Background: received multiple NSF awards for research on food security in developing countries; international focus, plus health and migration topics align with ICPSR collection policies
• Brief awards summary: food insecurity in refugee communities, and effects on mental health; ecological conditions’ effects on food security

Case Study #3

• Anthropology professor, Craig Hadley

• Considerations and concerns: collaborative nature of research, sensitivity of spatial data resolution, subjects’ informed consent at time of collection

• Outcome: interested, but had fair amount of documentation and related work to do
Conclusions

• ICPSR work in this area already is helpful – LEADS data, collection policies and priorities for acquiring data
• Our approach worked for identifying locally produced datasets of interest
• Alternatives for data sharing – our local Dataverse could be suitable when data are outside ICPSR collection priorities
• Time is the most valuable commodity in faculty professional lives – there is interest in sharing data, but need more support to find time to clean and document their data
Conclusions

• Importance of established working relationships between librarians and faculty

• Organizational structure poses a challenge to coordination
  • communication across units
  • time: finding the time for things that are not part of their core duties
Future Directions

• Needs of researchers informing our services
  • Support for data cleaning?
  • Support for data documentation?
• Alternatives for data sharing – reaching out to researchers about our local Dataverse when their data are outside ICPSR collection priorities
• NIH awards
Questions?

https://www.pinterest.ch/pin/497929302530049265/
References

• NSF awards list search: https://www.nsf.gov/awardsearch/advancedSearch.jsp
• NIH RePORTER awards database: https://projectreporter.nih.gov/reporter.cfm
• openICPSR: http://doi.org/10.3886/E100004V4 - "Evaluating Impact of the 2009 American Recovery and Reinvestment Act on Social Science"
• ICPSR collection-development policies: https://www.icpsr.umich.edu/icpsrweb/content/data management/policies/colldev.html (see "High-Priority Data" for areas of interest)