Assessing Data Repository Impact

Big Data Workshop
Measuring the Impact of Digital Repositories
NITRD
Arlington, Virginia
February 28-March 1, 2017
Introduction to ICPSR

- Founded in 1962 by 22 universities, now consortium of over 760 institutions world-wide
- Focus on social and behavioral science data, broadly defined
- Current holdings
  - 9,555 studies, comprised of 76,421 datasets and 257,843 files
  - 1200+ studies with online analysis capabilities
  - Bibliography of Data-related Literature with 70,741 citations
- Approximately 60,000 active MyData (“shopping cart”) accounts representing individuals in 40+ disciplines
- Thematic collections of data about addiction and HIV, aging, arts and culture, child care and early education, criminal justice, demography, health and medical care, and minorities
Assessing Impact: Currently

- Google Analytics
- Downloads by
  - Dataset/study
  - Collection/archive
  - Statistical package
  - Member institution (info available to ORs as well)
- SDA Usage statistics
  - sessions and datasets
- Learning Guide usage
- Webinar statistics
- Bibliography entries
- Numbers of
  - Data deposits
  - Summer Program participants
  - Disciplines and statuses represented in MyData accounts
- Social media followers, friends, retweets
- Feedback from ORs and others
Challenges with Current Methods

- No comparison metrics for clicks or downloads
  - Large numbers that grow, but how large should they be?
- Distinguishing impact on
  - Research
  - Policy
  - Training
- Bibliographic entries may not be comprehensive
  - Costly to locate
Community-based Measures of Impact

- Data Seal of Approval, since 2011
- Certification of infrastructure under ISO 16363 and CIPSEA
- Membership
  - International Council for Science World Data System
  - Data Documentation Initiative community
  - Research Data Alliance
Future Directions: Data Impact Factor

• Under development
• Similar to a citation or journal “impact factor”
• Based on the impact those data have on subsequent research
  – Requires and reinforces community standards for data citation
Qualitative Measures

• Data discoverability, usability into the future
  • Robust to technological change

• What are tradeoffs when considering cost of curation
  • E.g., relative to Institutional Repositories
  • Data easier to find due to metadata markup
  • Data easier to use with complete documentation

• Research transparency and replication

• New types and uses of data
  • linked data; harmonized or multiple datasets; naturally-occurring data
Qualitative Measures

• Continued diversification of users
  – New disciplines
  – Non-academic audiences (policy makers, journalists, community organizers, urban planners)

• Complementary training

• Outside adoption of ICPSR’s infrastructure architecture