Does Graduate Training in the Social Sciences Prepare Students for Data Management and Sharing?

Ashley Ebersole and Jai Holt
ICPSR University of Michigan
Graduate Training in Social Science

• The main objective of a graduate program in the social sciences
  • Producing quality scholars and researchers

• What skills should a good researcher have?

• Curriculum provided by these graduate programs
  • Research methods
  • Professional development
What’s missing?

• It is unclear whether graduate students are receiving training in data management or data sharing

• Social science researchers report not feeling satisfied with their training in data management (Johnke & Asher, 2012)
  • Primarily from on the job training
  • Minimal formal education
  • Unfamiliar with data services
Importance of Management and Sharing

- Propagation of scientific research
- Increasingly collaborative nature of research
- Issues of long term storage, loss of data, and reusability
- Importance of communicating research well in social sciences
“If you’re not sharing your data, you’re not doing good science”

- Data service professionals

- Ethical guidelines of many social science fields
  - Open access or release of data
  - Emphasis on making data usable by others

- Examining the ethical codes
  - Six fields of social science
Examining the ethical codes
Additional Needs for Data Sharing and Management


- Journals and funding sources such as the NIH, NSF (Tenopir, *et al.*, 2011)

- Difference between public access and sharing by request
Research Questions

• As a crucial aspect of science, ethically required by the governing bodies, are students being taught data sharing or management as they prepare to become independent researchers?
  • Coursework in ethics relating to the governing body of the field
  • Specific inclusion of data sharing or management information

• Are students leaving their social science graduate programs with adequate skills in data sharing and data management?
  • Knowledge of repositories or how to use them
  • Are faculty confident in their abilities?
Methods

• Two part digital survey
  • 1- Program demographics and content
    • Inclusion of Ethics coursework
    • Research methods course work
    • Specific inclusion of data sharing or data management
    • Self assessment of program effectiveness
  • 2- Use of repositories and interest in repository programming
    • Why are and why might students use repositories
Sample

• Six fields selected
  • Inclusion criteria: Social science, Research oriented; Major governing body with an available code of ethics
• Random selection from US schools in the gradschool.com list
  • Only programs with in-person courses included
  • Respondents for schools selected using program website
• 150 of schools surveyed
  • Private and public
  • All Carnegie classification levels included, as well as unclassified
  • Within all 50 states
Preliminary Results

- Trends from the initial sample
- Small group sizes limited cross field comparisons
- Some limitations in sampling
Good science is sharing data

- **Response Rates**
  - Total response rate – 18%
    - Total of 24 usable responses
  - Response rates for each field
    - Anthropology – 12%
    - Geography – 8%
    - History – 4%
    - Political Science – 28%
    - Psychology – 24%
    - Sociology – 20%

### Social Science Field Representation

<table>
<thead>
<tr>
<th>Program Field of Focus</th>
<th>Number of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociology</td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td></td>
</tr>
<tr>
<td>Political Science</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>2</td>
</tr>
<tr>
<td>Anthropology</td>
<td></td>
</tr>
</tbody>
</table>

12
Respondents

- Position of respondent
  - Program Coordinator (33.3%) or the Program Director (41.7%).
  - Other respondents

- Program size
  - Total number of students (n = 24) averaged 54.33 (sd = 75.61)
  - The number of full time faculty ranged from 2 to 35, with a median of 7.

- Degrees offered
  - 79.2% offer an MA
  - 20.8% offer an MS
  - 70.8% offer a PhD
Results: Research Methods

• 95.8% of programs reported a required research methods course
  • 70.8% (n = 22) also require students to complete a research project

• Statistical packages training
  • In 87.5% (n = 22) of programs a stats package is used in completing research
    • Most students use SPSS - 62.5%
    • STATA (50.0%), R (45.8%), SAS (29.2%)
  • Many programs allow students to choose which package they use (45.8%)
    • Only 16.7% require a specific package

• Adequacy of training in stats package use?
  • Reliance on undergraduate training or other experience
Results: Ethics in Research

• Do programs provide discipline specific ethics training?
  • 20.8% (n = 24) of programs include an ethics course
    • Psychology programs included ethics courses more frequently than any other programs
  • 12.8% report ethics being incorporated into another course
  • Of those which include an ethics course:
    • 80% included material from the governing body/ethical code
    • 100% reviewed research ethics
Results: Data Management

- Inclusion of Data Management in program
  - Course in Data Management
    - Only 1 program reported providing a course specifically in data management (4.2%)
    - 2 programs reported a plan to offer one
  - Of those that do not offer a course in data management
    - Four report not providing information about data management in another way
    - Eight report not knowing if the information is shared at all
  - 75% of programs report data management being included in Research Methods
Results: Data Sharing

• Within the Research Methods course
  • 37.5% report covering data sharing
  • 12.5% report only covering repository information

• Specific repository content or access
  • 45.8% report that this information is covered in coursework
  • 4.2% give students this info to review independently
  • 41.7% do not cover information about repositories

• Inclusion of information some other way
  • Assistantships
  • Advisors

![Data Sharing and Repository Inclusion in Research Methods](chart.png)
Results: Self assessment of program effectiveness

- Mean percentage of graduates working in research: 57.18% (n = 22, sd = 26.14)
  - 62.5% spending either half or a minority of time working on research

- Majority of programs report feeling confident students have adequate research skills (66.7%)
  - 70.9% report feeling moderately to extremely confident

Graduate Programs Level of Confidence in Student Research Skills

- Not at all
- Not very
- Slightly less
- Neither
- Moderately
- Very
- Extremely

Percentage of Programs
Results: Self assessment of program effectiveness

• Practical experience while in the program, on average
  • Publishing papers – 34% (alone); 41.6% (coauthor)
  • Presenting posters – 66.3%
  • Giving talks – 63.18%

• Half of programs report students using data repositories
  • For references, getting data – 27.3%
  • To share data – 4.5%
Results: Self assessment of program effectiveness

- Belief in graduates understanding of data sharing importance
  - 50% are unsure
  - 29.2% do not believe graduates have an understanding

- Belief in graduates data management skills
  - 41.7% believe graduates have good skills
  - 16.7% do not believe, 16.7% are unsure, and an additional 16.7% report mixed results
Results: Use of Repositories

- Most common reasons for currently using a repository:
  - Supplemental data
  - Developing a research question
  - Finding references or resources
Results: Use of Repositories

- Most frequent reasons why a student might use a repository:
  - Accessing supplemental data
  - Developing a research question
  - Finding resources or references
  - Verifying the originality of their ideas
Results: Interest in Repository Programming

- Most were unaware of any programs
  - 66.7% report no students attending any
- The majority of programs reported potential interest
  - 20.8% reported a definite interest
  - 62.5% reported maybe being interested
- Likelihood of recommending students participate

![Graph showing likelihood of recommending students to participate in Repository Programming]
Level of Interest in Data Management or Sharing Programs Ratings

Number of Programs

- Very Disinterested
- Mostly Disinterested
- Somewhat Disinterested
- Neither Interested nor Disinterested
- Somewhat Interested
- Mostly Interested
- Very Interested

Programs:
- In-person Seminars/Workshops
- In-Person Data Management Course
- Online Data Management Course
- Student Webinars
- Educator Webinars
- Website with Information and Resources
- Documents or Infographics
Perceived Level of Student Interest in Data Management or Sharing Programs

- In-person Seminars/Workshops
- Online Data Management Course
- Website with Information and Resources
- In-Person Data Management Course
- Webinars
- Documents or Infographics

Number of Programs

Very Disinterested | Mostly Disinterested | Somewhat Disinterested | Neither Interested nor Disinterested | Somewhat Interested | Mostly Interested | Very Interested

0 2 4 6 8 10 12

25
Conclusions

• Few programs include discipline specific ethics
• More programs than expected included information about data management
  • This information, when provided, was almost always within another course
• Training in data sharing or data repository use appears to be limited
• Few students appear to be using repositories in graduate school
  • Limited exposure to use and to sharing during training
• Despite what appears to be incomplete training, programs report high levels of confidence
  • More confident in data management than in data sharing
What can we do as data service professionals?

• Offer trainings and resources as much as we can
  • Results suggest a general interest in programs provided by repositories
    • Challenges of participation in current programs

  • Interest in other types of programs or items
    • In-person strategies
    • Online strategies
      • Coursework
      • Resources
What can we do as data service professionals?

• Connect with graduate students now
  • More inclined to connect with us in the future

• Prepares us to adapt and meet the changing needs of our consumers

• Novel research and new types of data
  • Help work out questions for storage and dissemination
Future Questions

• Project expansion
  • Including more social science subfields
  • Comparison with other disciplines
  • Surveying students – do they interpret the same way?

• How do we convince established researchers to share data?
Acknowledgements

• Jai Holt
• Kathryn Frania
• Amy Pienta
• Peter Granda
• ICPSR
References

