

New Incentive Infrastructure for Sharing Data and Other Research Outputs

Bommae Kim^{1,2}, PhD

CNI Spring 2017 Membership Meeting
April 4, 2017

¹ Research Methodologist, Federal Reserve Bank of Kansas City; bommae.kim@kc.frb.org

² CLIR Postdoctoral Fellow in Data Curation for the Sciences and Social Sciences (2016-2018)

The views expressed in this presentation are those of the speaker and do not necessarily reflect the position of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Outline

- Point of View
- Problems: No incentives for researchers to share research materials
- Solutions: Recognize each research material as an independent output
- Expected Effects
- Suggestions (Action Items)
- Discussion & Take-Home Messages

Point of View

How to motivate researchers to share research data and materials

What would be good selling points to researchers?

Social science

Research Materials (Data and More)

By-products of a research paper

- Data
- Software code to analyze data
- Software code to collect data
 - e.g., web scraping, matching multiple big datasets
- Other study materials to collect data
 - e.g., questionnaire, study stimuli

What people say about the benefits of data sharing

Research integrity (transparency and reproducibility)

Funding requirement

Re-use of resources (and being cited)

- Exact replication (no recognition)
 - To verify the study results (research integrity)
 - To use them in class (education)
- New research (great benefits for receiver, maybe-incentives for provider)
 - Using the same data (ONLY FOR meta-analysis and review)
 - Collecting new data using the same/modified tools (advancement of knowledge)

What people say about the benefits of data sharing

Research integrity (transparency and reproducibility)

Funding requirement

Re-use of resources (and being cited)

- Exact replication (no recognition)
- New research (great benefits for receiver, maybe-incentives for provider)

Data collecting tools are more reusable than data itself for new research

Researchers would like to have offshoots of their research

Problems

Top-down approach:

- Funder requirement

- Journal requirement

No incentives for individual researchers

Poor quality (low usability)

Researchers are not used to sharing data and other materials

Even when they do so, shared materials are of low quality

Solutions

Recognize each research material as an independent item

Separate authorships from the article

Credit “actual” authorship for each material

Expected Effects

Incentivizes and empowers unrecognized contributors

→ Increases responsibility, improves quality

Ingrains good research practices early in career

Enables flexible collaboration, facilitates interdisciplinary collaboration

Promotes research integrity (better data sharing)

Fosters research ethics (fair credit sharing)

Suggestions (Action Items)

Assign independent DOIs for each research material

Cultivate a new culture:

- Research materials can be recognized independently apart from the article
- RAs can be entitled to the authorship of research materials

Provide “matchmaking” service for collaboration

Discussion

Take-home messages

Other research materials can be more valuable than data

Research materials can have different authorships from the paper

RAs should be entitled to authorships of research materials

Bommae Kim

bommae.kim@kc.frb.org

Thank you

Appendix

Quality of Shared Data

Replication in Economic Research

“It is the policy of the American Economic Association to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication.” - AEA

- 13% of 54 replicated (Dewarld, Thursby, & Anderson, 1986)
- 22% of 62 replicated (McCullough, McGeary, & Harrison, 2006)
- 33%, 43% of 67 replicated (Chang & Li, 2015)

What does this tell us about the quality of data shared (due to journal requirement)?

Replication in Economic Research

# of articles	67 (100%)
Replicated	22 (33%)
No data or code	21 (31%)
Incorrect data or code*	9 (13%)
Data or software not available	8 (12%)
Needed the original authors' help*	7 (10%)
# of articles with data/code	38 (100%)
Problems with shared data/code*	16 (42%)

Chang & Li (2015)

67 papers (13 journals,
2008 - 2013)

Research Funding Status by Disciplines

Funding in Research

Among all research projects, how many of them are funded?

Among the funded research projects, how many of them are required to share data?

Funding in Research

Journal	Total	Financial Support				Corp	NONE
		NSF/NIH	Other	etc	School		
Material Science (<i>Advanced Materials</i>)	10	3 (30%)	7 (70%)				0 (0%)
<i>The New England Journal of Medicine</i>	14	7 (50%)				6 (43%)	1 (7%)
Computer Science (<i>ACM SIGKDD</i>)	40	5 (13%)	5 (13%)	4 (10%)	1 (3%)	23 (58%)	2 (5%)
Social Science (3 combined)	36	4 (11%)	7 (19%)		9 (25%)		16 (44%)

<i>Journal of Financial Economics</i>	10		2 (20%)		5 (50%)		3 (30%)
<i>Psychological Science</i>	11	4 (36%)	2 (18%)				5 (45%)
<i>American Journal of Political Science</i>	15		3 (20%)		4 (27%)		8 (53%)

END

References

Chang, A. C., & Li, P. (2015). Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not”. Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System. <http://dx.doi.org/10.17016/FEDS.2015.083>

Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. American Economic Review 76, 587-603.

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB Archive. Journal of Money, Credit and Banking 38, 1093-1107.