Data Sharing in the Academic Environment

National Academy of Medicine

Alexander Ommaya, DSc.
Senior Director, Scientific Affairs

November 3, 2016
THE DATA ARE MINE!

IVORY TOWER OF SCIENCE

SCENE FROM THE PAST?

Auke Herrema (sciencenode.org/feature/sharing-research-data-benefit-all-society.php)
Data Sharing Highlights

• National survey noted data requests honored increased and refusals declined between 2003 and 2013
• 100 random datasets selected from Gene Expression Omnibus were reused and cited 550 times over 5 year span
• For the industry sponsored CDSR repository, only 2% of proposals submitted involved reanalysis
• Several journals (inc. PLOS & Nature) require papers to include data availability statements and use standard data citations
• Primary reason for BioLINCC data requests was that collecting data of similar size not feasible

(# 3,4,5,7,8)
Data Sharing Highlights

• Measures for data sharing and data utilization have been proposed. Elsevier has committed to implementing research data metrics.

• A recent study noted (#6) that even with a strong data sharing policy, data sharing rates in BMJ are low.

• Funders are increasingly preferring or mandating open sharing of research data (NIH, NSF, Gordon and Betty Moore foundation, Bill and Melinda Gates foundation, Wellcome Trust, HHMI, PCORI)

(# 8,6)
Open sharing of research data - investigator benefits

• Engaging in open sharing of data may provide competitive advantage for research sponsorship
• Submitting data and documentation to repositories ensures preservation and accessibility
• Increases citation of work
• Increases visibility and opportunity for new collaborations
• Supports team science (increasingly part of promotion and tenure)

(# 2, 4)
Perceived barriers

• Inappropriate use due to lack of understanding research parameters
• Security and confidentiality
• Lack of acknowledgement and credit
• Reduced advantage when competing for grants
• Managing consent / de-identification
• How to work towards individual expressed consent for open sharing of data
Withholding and Sharing

• Leading reasons for *data withholding*
  • Not wanting to expend the effort to produce data, backup materials, and gain approvals
  • Protect ability of researcher, colleague, or postdoc to publish

• Leading reasons for *data sharing*
  • Formed collaborations that led to publications
  • Formed collaborations that led to grants

(# 3)
Lead time to share data

- Formatting data
- Describing scope of consent and data usage
- Preparing documentation, data dictionaries, etc.
- Obtaining required institutional approval
- More challenging when multiple institutions are involved

(#) 1
Lead time for accessing data

• Finding relevant and usable data

• Getting authorization to access restricted data (e.g. from dbGaP)

• Approval from supervisors and IT department

• More challenging when multiple institutions are involved (IP, data sharing, etc)

• Formatting accessed data (converting, standardizing, quality control, customizing, etc)

(# 1)
Institutions

• Libraries are often the POC for these issues
• Do provide instruction and linkage to public and private repositories such as Figshare, Dryad, Zenodo, Dataverse
• Meaningful sharing requires sufficient metadata, data dictionaries, annotations, protocols
• Opportunities for institutions to provide additional training – research sponsor requirements, data sharing tools, data formats, preparing documentation, etc.
References


