What we learned yesterday

- Statisticians bring expertise in design and uncertainty (at multiple levels – sample, measurement, analysis)
- We are called upon to translate and clarify meaning in our audience's language (communication)
- Good science requires good process control, ethical thinking, and transparency (by us and scientists)
- We need principles and systematic approaches for the data life cycle (data collection to preservation)

IOWA STATE UNIVERSITY

Office of the Vice President for Research

Open Science and the Need for Education about Data

Sarah Nusser

Professor, Statistics Department and
Center for Survey Statistics and Methodology

Vice President for Research

Iowa State University

ASA Chairs Workshop, July 2016

Premise of Inquiry

"Open inquiry is at the heart of the scientific enterprise"

***Science as an Open Enterprise Summary Report*, The Royal Society, 2012**

- Historically, this has involved:
  - Publication of theories, experimental and observational data, interpretations
  - Advancement of ideas through exchanges, colloquia and other venues
  - Self-correction through debate and scrutiny

Practice of Inquiry is Changing

- Centrality of printed page receding with digital technologies
- Increased emphasis on sharing data publicly
- Large-scale data collection and analysis is challenging traditional autonomy of individual researchers
- Internet provides a conduit for networks of scientists and public to collaborate and communicate

Open Science

Open data

[available, intelligible, assessable & useable]

combined with

Open access to scholarly publications

and

Effective communication of their contents

[among scholars and with public]

royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe-summary.pdf

It's not just making data available …

- Effective communication through more intelligent openness when sharing data
  - Understandable to those who wish to scrutinize data
  - Assessable for evaluating reliability of data and competence of data producers
  - Usable by others (scholars, public) for understanding and new discoveries

External Drivers

- Increased public interest in evaluating the credibility of scientific conclusions and underlying evidence
- Push for government accountability through public release of data (e.g., funding agencies)
- Hopes that open access will increase public trust in scholarship, new discoveries and business activity
- *Need for science and scholarship to adapt to the changing technological, social and political environment*

US Research Funding Agencies

- 2013 OSTP memo directed major research funding agencies to:
  - Ensure publications and data from research funding are shared
  - Protect privacy and confidentiality, proprietary information, national security
  - Balance value of long-term archival with costs and burden
  - Enable costs to be covered in grants

Data Sharing Practice in its Infancy

- Most data sharing occurs on a small scale (local, known collaborator), except for select disciplines
- Disciplinary cultures vary widely in their practices for preparing, sharing, curating data
- Agency guidance and repository systems emerging, but still underdeveloped and inconsistent across agencies
- Research sponsors have emphasized data for publications (data management plans), but there are many other data products from research

Preparing to share data:
the need to educate researchers & data scientists

- What is purpose of sharing? Audience? Value?
  - Validate publication OR for reuse, reproducibility
  - Scholars vs public
  - Value brought to science, scholars and public
- Given the purpose, what data should be shared?
- Should (part of) the data be protected?
  - Type of risk (e.g., confidentiality)
  - Risk of disclosure
- What options are available for protecting data?
  - Disclosure limitation methods (statistical, access)

Education for sharing data:
the need to train researchers and data scientists

- How do we effectively "communicate" data?
  - Importance of early planning and continuously updating documentation
  - Context of goals, methods, data manipulations, flaws
  - Standards for documenting data
  - Curation practices
- What are the options for storing data?
  - Data commons, federation of repositories
- Tools for proactive planning in research process?
  - Open Science Center, Research Data Alliance, …

IOWA STATE UNIVERSITY

Office of the Vice President for Research

Open Science and the Need for Education about Data

Sarah Nusser

Professor, Statistics Department and
Center for Survey Statistics and Methodology

Vice President for Research

Iowa State University

ASA Chairs Workshop, July 2016