QDR
QUALITATIVE DATA REPOSITORY

Data Under Constraint: Tools and Strategies for Facilitating Transparency

The third workshop in a series on "Developing and Implementing Data Policies: Conversations Between Journals and Data Repositories."

San Francisco, August 30, 2017

Colin Elman

QDR, Syracuse University

Agenda

Lunch (1:00pm – 1:30pm)

Introductions (1:30pm – 2:00pm)

Dataverse, Journals, and Sensitive Data (2:00 – 2:30pm), Gustavo Durand, Institute for Quantitative Social Science, Harvard University

Adapting Data Verification Workflows to Accommodate Restricted Replication Data (2:30 – 3:00pm), Thu-Mai Christian, Odum Institute, University of North Carolina

Break -- 3:00pm-3:15pm

QDR

Agenda

ICPSR's Restricted-use Data Management and Virtual Data Enclave (3:15 – 3:45), Justin Noble, (ICPSR)

Using Roper Center Data to Satisfy Transparency Requirements (3:45 – 4:15pm), Peter K. Enns, Roper Center

Break -- 4:15pm-4:30pm

Journal Editors Discussion Interface (JEDI) (4:30pm – 5:00pm) Colin Elman, Qualitative Data Repository (QDR)

Group Dinner for Workshop Participants (7:00pm), Le Colonial

QDR

Background

- Data Access and Research Transparency (DA-RT)
  - Data Access
  - Production Transparency (documentation)
  - Analytic Transparency (supplemental materials)

- Developments in political science part of much broader movement across the social sciences

QDR

Open Data and Transparency

BITSS

Berkeley Initiative for
Transparency in the Social Sciences

ipa
INNOVATIONS FOR POVERTY ACTION

Research Transparency

the WHITE HOUSE
PRESIDENT BARACK OBAMA

BRIEFING ROOM | ISSUES | THE ADMINISTRATION | 1600 PENN

HOME - BLOG

Increasing Access to the Results of Federally Funded Science

RESEARCH DATA SHARING WITHOUT BARRIERS

COS
CENTER FOR OPEN SCIENCE

About COS | Our Products | Our Services | Our Communities | Blog | Contact Us | Donate

OPEN, TRANSPARENT, AND REPRODUCIBLE SCIENCE
IS STRONGER SCIENCE.

RDA
RESEARCH DATA ALLIANCE

ABOUT RDA

DA-RT

DATA ACCESS & RESEARCH TRANSPARENCY

PLOS' New Data Policy: Public Access to Data

QDR

Drivers of Transparency

Data Availability

Journal Policies

The following policy applies to all PLOS journals, unless otherwise noted.

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal

http://journals.plos.org/plosone/s/data-availability

Insecure researchers aren't sharing their data

Norms / peer pressure

Posted by Andrew on 4 November 2011, 10:14 am

Jelte Wicherts writes:

> I thought you might be interested in reading this paper that is to appear this week in PLoS ONE.
>
> In it we [Wicherts, Marjan Bakker, and Dylan Molenaar] show that the willingness to share data from published psychological research is associated both with "the strength of the evidence" (against H0) and the prevalence of errors in the reporting of p-values.
>
> The issue of data archiving will likely be put on the agenda of granting bodies and the APA/APS because of what Diederik Stapel did.

I hate hate hate hate hate when people don't share their data. In fact, that's the subject of my very first column on ethics for Chance magazine. I have a

http://andrewgelman.com/2011/11/04/insecure-researchers-arent-sharing-their-data/

Dissemination and Sharing of Research Results

Funders

NSF DATA SHARING POLICY

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF DATA MANAGEMENT PLAN REQUIREMENTS

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

https://www.nsf.gov/bfa/dias/policy/dmp.jsp

Professional Organizations

6. Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.

6.1 Data access: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they

American Political Science Association

A Guide to Professional Ethics in Political Science

http://www.apsanet.org/portals/54/Files/Publications/APSAEthicsGuide2012.pdf

QDR

Concerns and Cautions

- Epistemic
- Costs and Logistics
- Ethical and Legal ➔ data under constraint

QDR

Data Under Constraint

- **Ethical** – human subjects concerns. Almost invariably involves institutional component, because of IRB requirements. Text of informed consent.
- **Proprietary** – data that author uses under a license granted by an owner, but to which other users do not have access without obtaining their own license.
- **Copyright** - an intellectual property right assigned automatically to the creators of “original works of authorship” (title 17, U.S. Code), which prevents unauthorized copying and publishing of an original product.

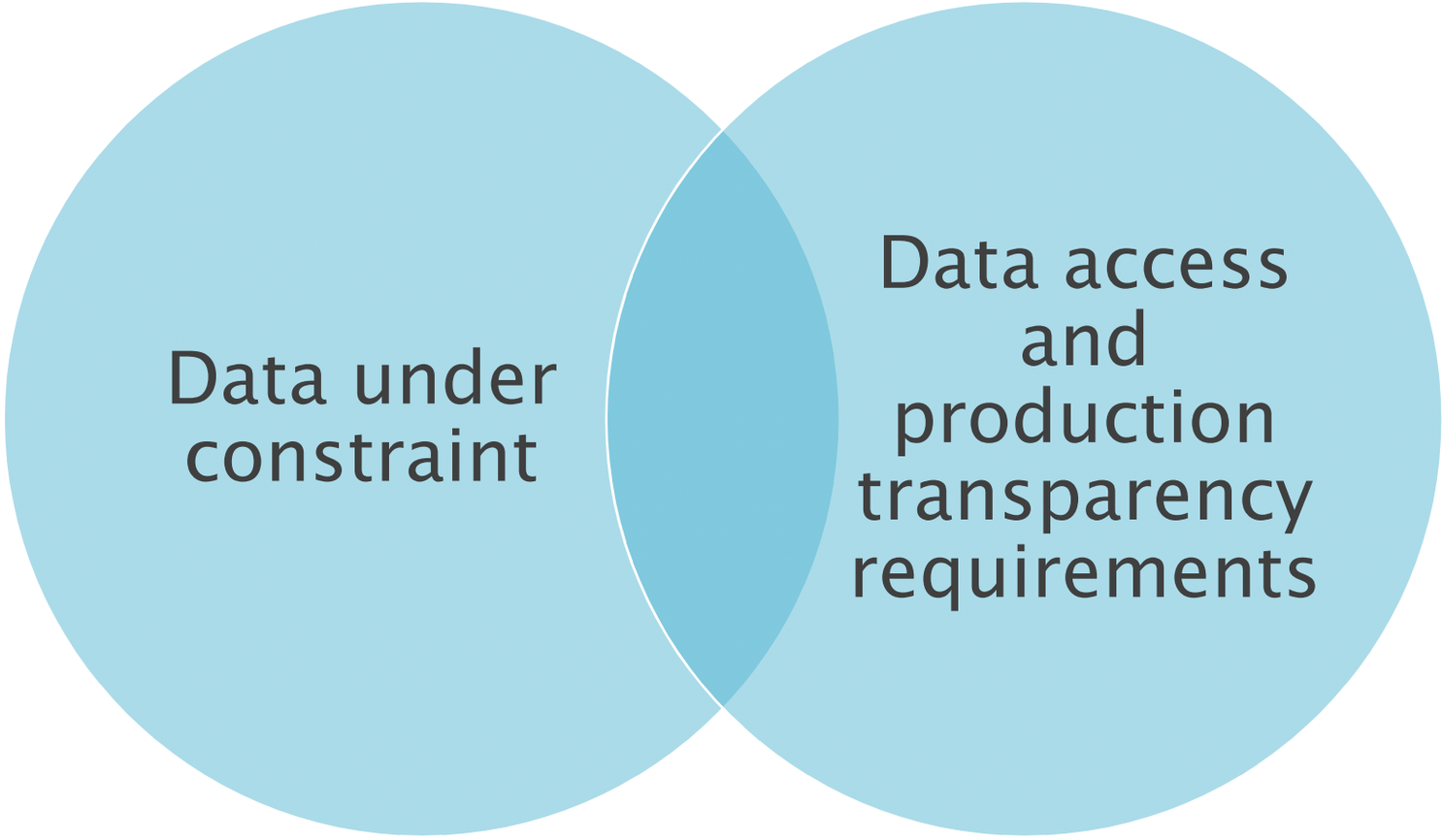
QDR

Not “all or nothing”

- Not a simple dichotomy between data that are toxic and shareable.
- Researchers have some agency in making choices that can render potentially problematic data shareable.
- Repositories provide infrastructure that can expand available choices, and empower researchers to satisfy ethical and legal constraints and share their data.

QDR

Expanding the middle ground



QDR

Five Safes framework

- **Safe projects:** Is this use of the data appropriate? Looks at moral, ethical and legal questions about the use of the data.
- **Safe people:** Can the researchers be trusted to use it in an appropriate manner? Asks whether they have the knowledge, skills, and incentives to store and use the data.
- **Safe data:** Is there a disclosure risk in the data itself? Is there a potential for re-identification?

Source: Tanvi Desai, Felix Ritchie, and Richard Welpton, Five Safes: designing data access for research, Economics Working Paper Series 1601, 2016, University of the West of England

QDR

Five Safes

- **Safe places/settings:** Does the access facility limit unauthorized use? Ranges from physical supervision through unfettered online access.
- **Safe outputs:** Are the statistical results non-disclosive? Residual risk in publication from sensitive data.

Source: Tanvi Desai, Felix Ritchie, and Richard Welpton, Five Safes: designing data access for research, Economics Working Paper Series 1601, 2016, University of the West of England

QDR

Five Safes

- Measures, not states.
- Generic framework that is intended to apply broadly, public and private, quantitative and qualitative.
- Several and joint contribution to assess data strategies
  - *several* – assessed individually and independently
  - *joint* – work conjuncturally, and together produce an overall level of risk.

QDR

Five Safes are changeable

- **Safe data**: risk of re-identification can be reduced by modifying the data.
- **Safe settings/places**: Physical isolation and supervision, and secure technologies (e.g. remote access, virtual enclaves)
- **Safe people**: Certification, training and data use agreements; bonds
- Technology is expanding the ways that risk can be reduced. Expanding the middle ground.

QDR

How can repositories expand the middle ground?

- Engage with IRB community, increase institutional space for compromise position (e.g. September workshop)
  - Create awareness for data sharing norms
  - Include in informed consent
- Provide support for different technical solutions that expand middle ground. For example:
  - Host de-identified data
  - Secure storage and access controls
  - Researcher credentialing

QDR

Workshop goals

- Spark a conversation about relationship between clashing mandates.
- Familiarize journal editors with recent and forthcoming developments at repositories to host data that are under constraint.
- Discuss whether (and if so how) to include language in authors guidelines allowing for restricted data that are protected using one or more of the mechanisms
- Introduce JEDI, opportunity for continuing this and other conversations.

QDR