Human Subjects Data in an Open Access Repository: Considerations and Challenges

Erin Clary and Debra Fagan

Dryad Digital Repository

Curated resource that makes the data underlying scientific publications discoverable, freely reusable (CC0), and citable. Dryad provides a general-purpose home for a wide diversity of datatypes.

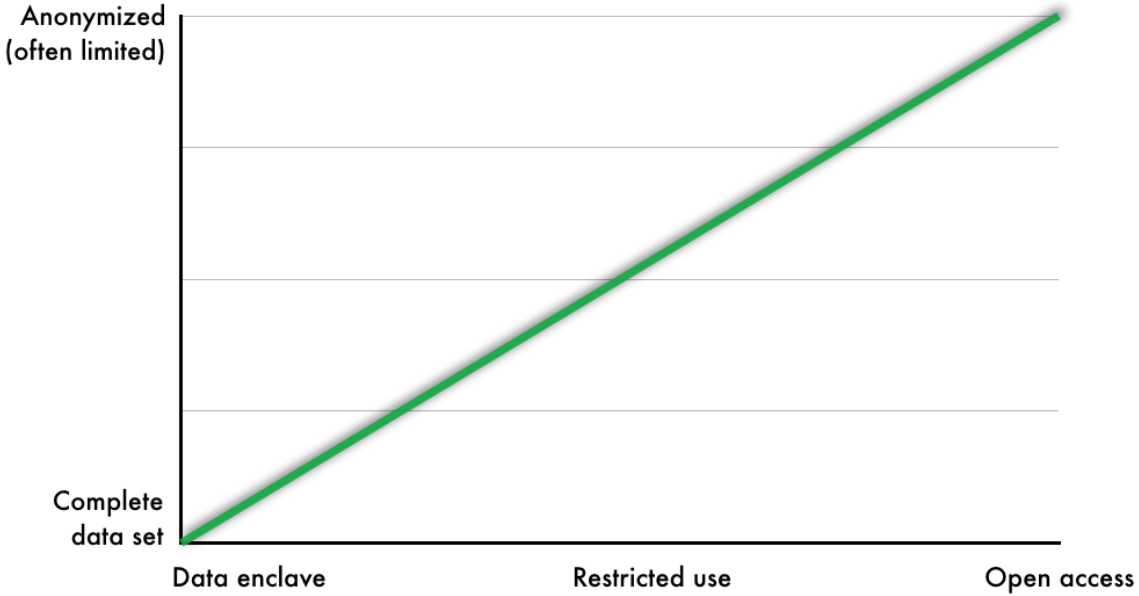
Our mission is to provide the infrastructure for, and promote the re-use of, data underlying the scholarly literature.

Can human subjects data be open access?

YES! But only if appropriately anonymized

- Direct identifiers
  - e.g., name or SSN
  - Remove all
- Indirect identifiers
  - e.g., age or gender
  - When considered together, may make unique cases visible
  - Number of indirect identifiers must be limited

Can human subjects data be open access?



Further Discussion...

Please join me for further discussion in Windflower AB

Recent protections in human subjects research
Why share human subjects research?
How to safely share human subjects data
Human subject submissions to Dryad–Preparing for public release
Do the benefits outweigh the risks?

Human Subjects Data in an Open Access Repository: Considerations and Challenges

Erin Clary and Debra Fagan

Background

- Dryad
  - Partner with journals
  - Required terms for archiving are CC0
- Journals/Societies/Publishers
  - Increasing number of journals are requesting or requiring that authors share the data underlying their publications
- Funders
  - Mandates to share research products

Outline

Recent protections for human subjects
Why share human subjects data?
Options for sharing
Examples from Dryad
Do benefits outweigh risks? - Group discussion

Recent protections - International

Nuremberg Code (1947)

voluntary informed consent

Declaration of Geneva (1948)

Helsinki Declaration (1964, last revised 2013)

"It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects" (WMA, 2013)

WHO Operational Guidelines for Ethics Committees (2000)

Protections for human subjects - national

Public Health Service - mandates IRB (1966)

"Safeguarding the rights and welfare of human subjects involved in research support by PHS grants is the responsibility of the institution to which the grant is awarded"

Tuskegee Syphilis study outed (1972)

National Research Act (1974)

Belmont Report (1979)

HHS Code of Federal Regulations Title 45 part 46 - protection of human subjects (expanded 1979)

"Common Rule" (45 CFR 46, subpart A) (1991)

Protections for human subjects - publishers

The International Committee of Medical Journal Editors (ICMJE)

recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals

The Committee on Publication Ethics (COPE)

"Code of Conduct and Best Practice Guidelines" include guidelines on protecting individual data and encouraging ethical research

DRYAD

Why share human subjects data?

"Researchers, authors, sponsors, editors and publishers all have ethical obligations with regard to the publication and dissemination of the results of research. Researchers have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports ... Negative and inconclusive as well as positive results must be published or otherwise made publicly available." (World Medical Association, Declaration of Helsinki, 2013)

Why share human subjects data? Benefits

Transparency in Science
Reduce duplication in effort
Allow exploration of topics beyond those originally envisioned
Educational use
Credit to researchers
Inform clinical practice and public health interventions

How to safely share human subjects data

Access control

- Data remain in full form, with restrictions on use in place
- Data will often be more complete, and therefore may be more useful

Open Access

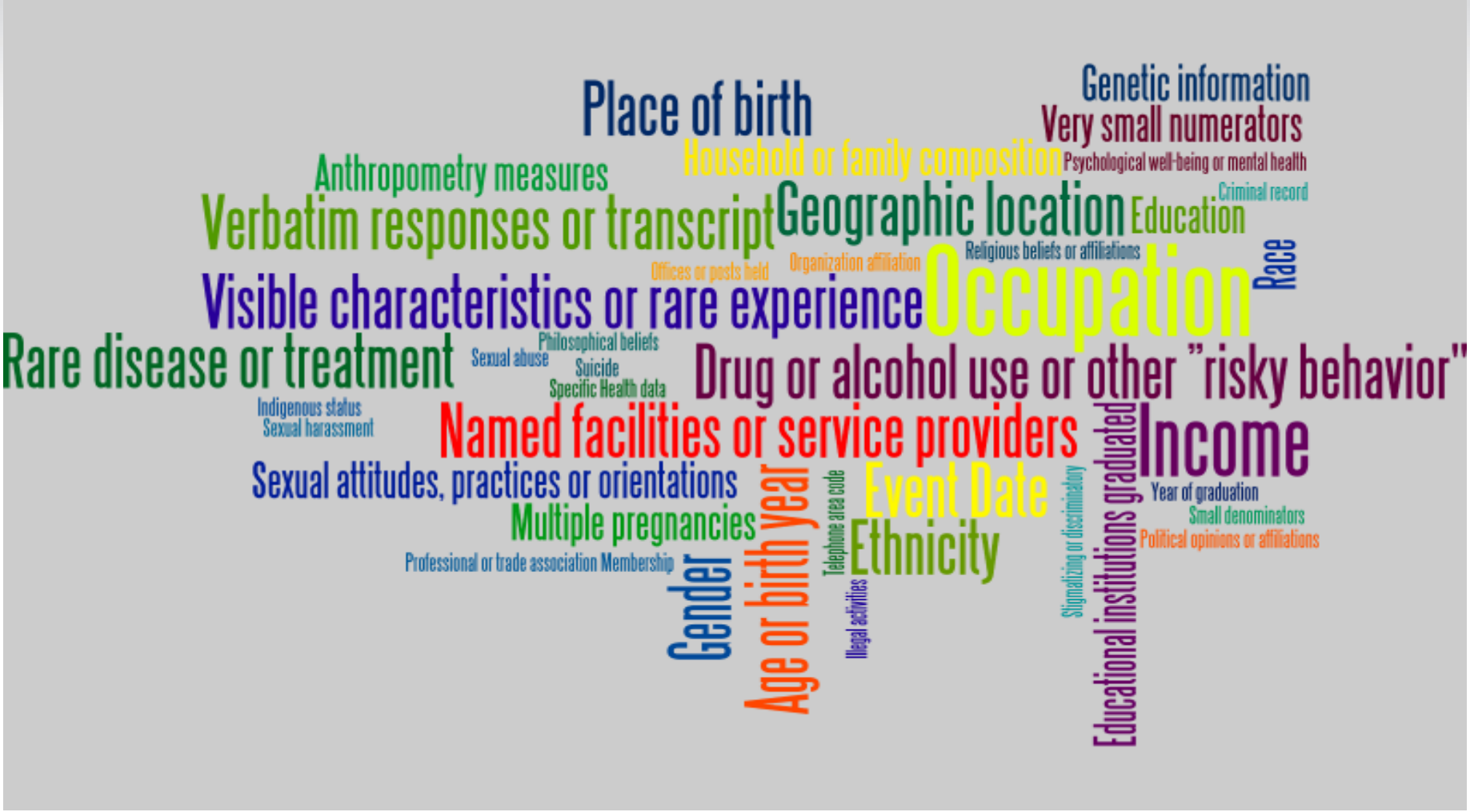
- Anonymize data set for public release
- More users may have access
- Variables included may be limited, so data may not be as useful

Identifiers

Direct identifiers are “variables that point explicitly to particular individuals or units” (ICPSR, p. 37)

- e.g., Name, initials, DOB, address, telephone, fax, or other contact information, e-mail address, Web or IP addresses, SSN, vehicle identifiers, medical device identifiers, biometric data (fingerprints, retina scans, palm lines, DNA), facial photograph, audiotapes, names of relatives, dates related to an individual, spatial location

Indirect identifiers are variables that when combined may “make unique cases visible” (ICPSR, p. 37)


Place of birth
Genetic information
Very small numerators
Household or family composition
Psychological well-being or mental health
Anthropometry measures
Criminal record
Verbatim responses or transcript
Geographic location
Education
Religious beliefs or affiliations
Offices or posts held
Organization affiliation
Occupation
Race
Visible characteristics or rare experience
Rare disease or treatment
Philosophical beliefs
Sexual abuse
Suicide
Specific Health data
Drug or alcohol use or other "risky behavior"
Indigenous status
Sexual harassment
Named facilities or service providers
Income
Sexual attitudes, practices or orientations
Event Date
Year of graduation
Multiple pregnancies
Small denominators
Ethnicity
Political opinions or affiliations
Professional or trade association Membership
Gender
Age or birth year
Telephone area code
Illegal activities
Stigmatizing or discriminatory
Educational institutions graduated

Indirect identifiers are tricky

"Data identifiability can be characterized as being on a continuum, in which the division between degrees of 'identifiability' are not always clear-cut. Even a dataset without direct identifiers may present a risk of indirectly identifying data subjects if the dataset contains sufficient information about the individuals concerned." (Canadian Institutes of Health Research, p. 29)

Indirect identifiers are tricky

Guidance from The ICPSR Guide to social science data preparation and archiving: Best practice throughout the data life cycle, 5th ed. and ANDS Guide to Publishing and Sharing Sensitive Data.

To limit risk, indirect identifiers must be removed or masked

Top or bottom code, collapse variables, round values, cell suppression, combine variables, sample, swap, disturb

Human subjects submissions to Dryad

- 3.5 month study to track human subjects submissions in Dryad
  - Data collected between Feb 20, 2015 and June 5, 2015
  - Packages associated with human subjects research were identified based on title/abstract
- Each package identified was tracked through the workflow
  - Number of times returned
  - Was it ultimately archived?

RESULTS

78 packages identified (approximately 5-7% of total submissions to Dryad)

33 packages returned for identifiable human subjects data (42.31%)

- 22 packages returned once (13 ultimately archived)
- 9 packages returned twice (8 ultimately archived)
- 1 package returned 3 times (ultimately archived)
- 1 package returned 4 times (ultimately archived)

62 of the human subjects submissions were ultimately archived (79.49%)

Our Challenges

- Time consuming
  - multiple curators to consult and/or multiple rounds of revision
- Diversity of data types from multiple disciplines
- Often assisting at the tail end of the process
  - Researcher may not have planned for public release, only archiving at journal request
  - Exceptions to policy can be made, but this is not always clear to the researcher
- Differences in International guidelines/requirements

Bigger picture challenges

Raw data may pose risk to privacy, but de-identified datasets may limit the benefits of data sharing

Need for clear documentation

Infrastructure and other technical support

Workforce

Sustainability / funding

For Discussion

Does the benefit of sharing outweigh the risks? If yes, what can we do to help mitigate risk? What should be considered in assessing risk/benefit.

Where should data be stored? Who should take responsibility?
(Institutional repository, researcher, publisher, funder?)

How to engage greater community?

"No stakeholder can single-handedly create a clinical trial ecosystem in which sharing data is expected and the risks of sharing are minimized. But all stakeholders have crucial roles and responsibilities in creating a culture of responsible sharing of clinical trial data and in providing effective incentives for such sharing." (Institute of Medicine, p. 4)

For Discussion

If benefits of sharing publicly do outweigh the risks, then what can we do to educate both researchers and information professionals? Where do we start?

Resources for guidance

The Inter-University Consortium for Political and Social Research (ICPSR). Guide to social science data preparation and archiving: Best practice throughout the data life cycle, 5th ed. Ann Arbor, MI: Institute for Social Research, University of Michigan, 2012. http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf.

Olesen, Sarah. ANDS Guide to Publishing and Sharing Sensitive Data. Melbourne, Australia: Australian National Data Service, 2014. http://ands.org.au/guides/sensitivedata.pdf.

National Institutes of Health. Human Subjects Protections Training. http://grants.nih.gov/grants/policy/hs/training.htm

Institute of Medicine (IOM). 2015. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington, DC: The National Academies Press.

References

Beecher, Henry K. “Ethics in Clinical Research”. June 16, 1966. The NEJM 274(24): 1354-1360. Retrieved from http://www.hhs.gov/ohrp/archive/documents/

Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research”
BMJ Open. “Frequently Asked Questions.” Accessed January 10, 2016.
http://bmjopen.bmj.com/site/about/faqs.xhtml

Canadian Institutes of Health Research. CIHR Best Practices for Protecting Privacy in Health Research. Ottawa, Ontario: Public Works and Government Services Canada, 2005. http://www.cihr-irsc.gc.ca/e/documents/et_pbp_nov05_sept2005_e.pdf.

COPE (Committee on Publication Ethics) - CODE OF CONDUCT AND BEST PRACTICE GUIDELINES
http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf

Creative Commons. “CC0 1.0 Universal (CC0 1.0): Public Domain Dedication.” Accessed November 20, 2015. https://creativecommons.org/publicdomain/zero/1.0/.

References

Department of Health, Education, and Welfare. “The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.” (1979, April 18). The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Accessed June 2015. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

“Dryad.” Dryad Digital Repository. Last modified November 19, 2015. https://datadryad.org.

Dryad Digital Repository (2013, August 22) Dryad Terms of Service. Retrieved from http://www.datadryad.org/pages/policies

European Parliament, Council of the European Union. “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.” EURLex. Accessed November 20, 2015. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046.

References

Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. “Preparing raw clinical data for publication: Guidance for journal editors, authors, and peer reviewers.” (2010). BMJ 340: c181. http://dx.doi.org/10.1136/bmj.c181

ICPSR. “Preserving Respondent Confidentiality.” ICPSR. Accessed November 20, 2015. https://www.icpsr.umich.edu/icpsrweb/content/deposit/confidentiality.html.

The Inter-University Consortium for Political and Social Research (ICPSR). Guide to social science data preparation and archiving: Best practice throughout the data life cycle, 5th ed. Ann Arbor, MI: Institute for Social Research, University of Michigan, 2012. http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf.

International Committee of Medical Journal Editors (ICMJE) - Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (2014) http://www.icmje.org/recommendations/

References

Institute of Medicine (IOM). 2015. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington, DC: The National Academies Press.

National Institutes of Health. Human Subjects Protections Training. http://grants.nih.gov/grants/policy/hs/training.htm

Office of the Surgeon General, Public Health Service. “Surgeon General’s Directives on Human Experimentation.” (1966) (Note, approved 1/24/67). Accessed January 5, 2016. https://history.nih.gov/research/downloads/Surgeongeneraldirective1966.pdf

Olesen, Sarah. ANDS Guide to Publishing and Sharing Sensitive Data. Melbourne, Australia: Australian National Data Service, 2014. http://ands.org.au/guides/sensitivedata.pdf.

N/A. Permissible Medical Experiments. Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10: Nuremberg October 1946–April 1949. Washington: U.S. Government Printing Office (n.d.), vol. 2, pp. 181-182. Text accessed June 2015 http://www.hhs.gov/ohrp/archive/nurcode.html

References

PLOS ONE. “Guidelines for Specific Study Types: Human Subjects Research.” Submission Guidelines. Accessed November 20, 2015. http://journals.plos.org/plosone/s/submission-guidelines#loc-guidelines-for-specific-study-types.

Sparks, Joel. “Timeline of Laws Related to the Protection of Human Subjects”. NOTE - last updated (2015, September 25). First published 2 February 2005. Compiled in 2002. Accessed January 5, 2016. https://history.nih.gov/about/timelines_laws_human.html

The Royal Society. “Data Sharing and Mining: Open Data Policy.” Publishing Ethics and Policies. Accessed November 20, 2015. https://royalsociety.org/journals/ethics-policies/data-sharing-mining/.

United States Department of Health and Human Serivces. “Title 45, Public Welfare, Part 46, Protection of Human Subjects” (2009, January 15). Accessed January 5, 2016. http://www.hhs.gov/ohrp/policy/ohrpregulations.pdf

References

United States Department of Health and Human Services, Office for Civil Rights. “45 CFR Parts 160 and 164: Standards for Privacy of Individually Identifiable Health Information. Final Rule.” Federal Register 67, no.157 (2002, August 14). DOCID:fr14au02-32. http://www.gpo.gov/fdsys/pkg/FR-2002-08-14/pdf/02-20554.pdf.

United States Department of Health and Human Services. “Frequently Asked Questions: Data Sharing.” National Institutes of Health, Office of Extramural Research. Revised February 16, 2004. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm.

United States Department of Health and Human Services. “Health Information Privacy: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA)
Privacy Rule.” Office of Civil Rights. Accessed November 20, 2015. http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html.

References

United States Department of Health and Human Services. “U.S. Public Health Service Syphilis Study at Tuskegee” (2013, December 30). Centers for Disease Control and Prevention. Accessed June 2015. http://www.cdc.gov/tuskegee/index.html

Van den Eynden, Veerle, Louise Corti, Matthew Woollard, Libby Bishop, and Laurence Horton. Managing and Sharing Data: Best Practice for Researchers, 3rd ed. Essex, UK: UK Data Archive, 2011. http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

World Health Organization (WHO). “Operational Guidelines for Ethics Committees That Review Biomedical Research” TDR/PRD/ETHICS/2000.1 (2000). http://www.who.int/tdr/publications/documents/ethics.pdf

World Medical Association (WMA). WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Fortaleza, Brazil: 64th WMA General Assembly, 2013. http://www.wma.net/en/30publications/10policies/b3/.