Sharing of Data: Human Subjects Issues and Data Management Plans

Natasha Warner
University of Arizona

Overall topic

- Protecting human subjects (source of data)
- While sharing data with other researchers
- Archiving it (data loss, accessibility)
- How the system winds up working

- Caveat: truly risky research vs. no-risk research

Interested parties

- Researchers
- IRBs/Human Subjects Protection Offices
- Funding Agencies
- Subjects themselves
- Communities/Tribes

Motivations/Goals: Researchers

- Sharing at least results, maybe data, publicly for the benefit of science
- Or for the benefit of their own career
- Not sharing data to keep the ideas and publications for oneself (getting scooped, publishing the most out of one's data, industry and applied research)
- Not sharing data because of lack of organization (filenaming, labeling, etc.)

Motivations/Goals: Human Subjects Offices

- Basic motivation: keeping data as private as possible so there can't be any risk
- Federal regulations often unclear
- PI is source of info about risk, but not objective
- Data that's completely private, or destroyed, can't possibly pose a risk, so err on the side of caution
- Increasing future data collection not a problem
- Results in nonsensical assumptions about risk

Motivations/Goals: Funding Agencies

- Goal of data sharing
- (Not the same as dissemination by publication)
- Maximize impact of funding
- At least for NSF, maybe not for applied funding, especially defense

Motivations/Goals: Subjects (Individuals)

- Often don't care about data sharing, only goal is getting the extra credit
- Students from Intro. classes: usually avoid sensitive content even during open conversation recording (but not always!)
- We usually play only short clips without obvious names

Others I'm not going to play…

- What if we do put the whole corpus on the web or at LDC etc. (including full personal conversations, not short clips)?

Riskier situations: Different goals

- Speaker discusses sexual orientation, medical information, union issues, illegal activity, etc.
- Researcher leads speaker to discuss those
- Speaker gossips about local people (Tribal government, boss, etc.)
- Study of gay people who aren't out
- Native American tribes that control members' participation in research or control sharing of the language
- Oppressed minority group: dangerous for ethnicity to be known

Summary of Conflicting Motivations

- Researcher: Share results, maybe share data
- Human Subjects Office: Don’t share anything
- Funding agency: Share raw data
- Speakers: Usually don’t care, except riskier situations
- So what happens?

How IRB rules induce PIs not to ask permission

- It IS possible to get permission for a lot (risky research, sharing of non-risky data) at most universities
  - if risks and sharing are clearly described
  - and subjects give written consent to them
  - PIs often assume they won't be able to get permission, so don't ask

Why PIs don't ask permission

- Picky questions on standardized forms (e.g. 'What is your plan for continuing data collection if subjects become incarcerated during the study')
- Presuppositions of badly written forms (e.g. 'When will data be destroyed')
- Not realizing "N/A" or "Data will not be destroyed because…" is acceptable"
- Fear/frustration: students, senior researchers

Possible Outcomes for data sharing

1. Researchers learn how to obtain permission -> share and archive raw data with not too much trouble
2. Researchers assume Human Subjects Office won't allow sharing -> don't attempt to share data
3. Human Subjects Offices do forbid data sharing -> no sharing

NSF Data Management Plans

- DMP requirement new, not well understood
- Requires a promise/plan/timeline to share raw data with other qualified researchers (not the public)
- Can be through a public archive, a web page, or “email the PI”
- Exceptions (e.g. PI does not own rights to data)
- Archiving and back-up requirement

Possible impacts of DMP

- DMP requirement encourages greater data sharing, discourages data destruction
- (Sharing has always been required)
- PIs may be able to use NSF panel summary as a tool with Human Subjects Offices
- DMP requirement pushes against the more careful Human Subjects Offices

The future and the ANPRM for the Common Rule

- US Federal Gov't considering complete re-working of human subjects regulations
- Especially strong impact on low-risk/no-risk behavioral research
- 1000+ comments submitted
- Will take a few years to find out what happens and to see how it will be implemented
- Implementation is bound to be variable

ANPRM as proposed

- Good:
  - ✧No-risk research: short form NOTIFYING Hum. Sub. Office of research, no real review
  - ✧PI determines whether it's "no-risk" (!)
  - ✧Consent simpler, maybe just oral
  - ✧Collaborative research across U.S. universities approved at only one university (data sharing simpler)
  - Bad:
  - ✧HIPAA regulations for all hum. sub. research from any US university

Conclusions

- Sharing of non-sensitive speech data is probably more possible now than many researchers realize
- DMP requirement may lead to greater sharing
- ANPRM could make data sharing much easier, but we don’t know yet.