Resource and Data Sharing
(or Why We are Never All Set)

Greg Simon
The privacy of participants should be safeguarded.
Data should be made as widely and freely available as possible.
Data should be shared no later than the acceptance for publication of the main study findings.
Initial investigators may benefit from first and continuing use of data, but not from prolonged exclusive use.
Original Mandate (from first RFA):

“The HCS Research Collaboratory program encourages sharing of study data from Demonstration Projects in a timely manner with appropriate privacy and confidentiality protections to facilitate further research, reuse of data, and replication. Thus the HCS Research Collaboratory program expects grantees to implement a Resources and Data Sharing Plan consistent with achieving these program goals.”
Why share resources and data?

- Transparency: People can see everything we did
- Reproducibility: People can re-create everything we did
- Secondary use: People can ask and answer questions we didn’t think of.
Shrinking realm of secondary use

• All health system data? – No
• All health system data regarding study participants? – No
• Final analytic variables regarding study participants? - Maybe
Barriers to secondary use

- Patient privacy
- Health system privacy
- HIPAA “minimum necessary information” requirement
- Scientific quality and integrity
Our Policy Summarized

• Investigators will each share, at a minimum, a final research data set upon which the accepted primary publication is based.
• Sharing data may require precautions to protect specific interests of collaborating health care systems.
• Precautions may include data sharing in more supervised or restricted settings, such as access to researchers who agree to limited pre-approved research goals.
• Investigators will choose the least restrictive method for sharing of research data that provides appropriate protection for participant privacy, health system privacy, and scientific integrity.
• Investigators will work with NIH to implement this data sharing policy, to ensure the appropriate administrative processes and technical infrastructure are in place.
Three levels of data

- What we can touch: original health system data
- What we can use: limited by consent or waiver
- What we can share: further limited by regulation (and possibly health system policy)

You can’t share what you never had!
Three levels of data

• What we can touch: original health system data
• What we can use: limited by consent or waiver
• What we can share: further limited by regulation (and possibly health system policy)

You can’t share what you never had!
Alternative structures and tools (from least to most restrictive)

- Private archive – send data to approved users
  - For any purpose
  - For approved purposes only (e.g. governed by DUA)
- Public enclave – any user may send questions to data
- Private enclave – some users may send questions to data
  - Automatically executed
  - Moderated (either queries or results)

It’s like real life: The more you want to control things, the more it’s going to cost you!
Three levels of data – applied to SPOT

- What we can touch: all original health system data
- What we can use: analytic dataset (failure-time structure) with outcome and short list of covariates (sex, age group, baseline item 9 score, site, race/ethnicity, primary care vs. MH specialty)
- What we can share: analytic dataset less site variable

Privacy risk is very low – but so is value for secondary analyses