Objectives

• Identify methods for tribal nations to acquire data from researchers and store data in a way that promotes secondary use.
• Discuss how tribal nations can benefit from secondary data analysis.
Preserving and Using Research Data

• Data stewardship
  – Different from ownership
  – Caretaking of the data
    • Accessibility
    • Preservation
  – Allows for potential future use of the data
Preserving and Using Research Data

• National data and surveys may lack data for American Indian populations

• Individual research studies might have more specific information

• Data from completed research studies may be used for additional analysis
  – Preliminary data for other research questions
  – Answer important questions for the tribe
Preserving and Using Research Data

• Obtaining accessible data
  – How should data be returned?
  – What format is the most useful for future use?

• Storing the data
  – How should data be preserved?
  – How can the data be kept safe?

• Using the data
  – What data can be used?
  – How can the data be used?
DATA TERMINOLOGY
Research Data

• Laboratory Notebooks
• Completed Surveys
• Summarized Tables and Statistics
• Research Findings or Published Results
• Laboratory Specimens
• Raw Data
Research Data

• The NIH defines final research data as: Recorded factual material commonly accepted in the scientific community as necessary to document and support research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based.
Research Data

Research Project

- Research Data
- Aggregate Data
- Other
Tribal Research Data

• Data Stewardship: Responsible for taking care of the data – ensuring quality of stored data
  – What will be done with the data?
  – Who will be responsible for collecting, storing, and analyzing the data?
Cataloging Research Studies

• Stores information about specific studies
• Can include:
  – Principal Investigator
  – Subject and Keywords
  – Relevant IRB forms and protocols
  – Dates
  – Other relevant information
• Usually this does not include research data
Collecting Research Data

• What data should be returned?
• Where will the data be stored?
• What will be done with the data?
Research Data: Terminology

• Raw (Primary) Data: Data that has been collected from a source but has not been subjected to processing or any other manipulation.

• De-Identified Data: Data that has been processed to prevent a person’s identity from being connected to information.
Research Data: Databases

- The primary method of storing data is in a database
- Data is organized into rows and columns

<table>
<thead>
<tr>
<th>IDNumber</th>
<th>FirstName</th>
<th>LastName</th>
<th>Address</th>
<th>City</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>142177</td>
<td>Dennis</td>
<td>Smith</td>
<td>101 W. 1st St</td>
<td>Sioux Falls</td>
<td>SD</td>
</tr>
<tr>
<td>145765</td>
<td>Joe</td>
<td>Anderson</td>
<td>431 S. Cliff Ave.</td>
<td>Sioux Falls</td>
<td>SD</td>
</tr>
</tbody>
</table>

Fields (Each column, blue boxes are just three of the six fields included)

Records (Each row, red boxes contain all information about a subject or person)
DATA RETURN
Data Return

• Why is data return important?
  – Tribal research
  – Education
  – Benefit

• Inform researchers early in the process
  – What exactly should be returned?
  – What format should it be returned in?
  – When should it be returned?
Data Return

• Success Story: Data Return Form
  – Sent to researchers upon completion of study
  – Items requested on a single flash drive:
    • Data folder containing:
      – Data in a universal format
      – Codebook and Instruments
    • Grant and relevant IRB forms
    • Newsletter and other forms of dissemination of results
Data Return

• Data Folder
  – Request to contain de-identified data in a universal format
    • Not platform or software dependent
    • .txt or .csv
    • Software specific (SPSS, Excel, SAS, etc.) files could also be included
Data Return

• Codebook and Instrument folder
  – A codebook is a descriptive file to explain the variables in the data set
  – A blank copy of survey instrument that was used
DATA STORAGE
Data Storage

• Data Access
• Security
• Location
  – Cloud
  – Server
  – Local Storage
• Procedures
  – File Naming
  – Backup/Recovery
Data Storage: Data Access

• Best Practices:
  – Data should be stored so that it can be accessed, but access should be limited
  – More than one person should have access
Data Storage: Security

• Password Protection
  – Store passwords in secure location
• Email is not secure – even to yourself
• Use firewall and anti-virus software to protect against cyber attacks
• Restrict access to computers, services, or locations containing data
Data Storage: Location

- Cloud
  - External company maintains server
  - Data is stored for a monthly fee
- Server
  - IT department usually maintains on-site
  - Log in required
- Local Storage
  - Flash Drives, Hard Drives, CDs, DVDs, etc.
# Data Storage: Location

<table>
<thead>
<tr>
<th>Storage Type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud</td>
<td>• Easy to use</td>
<td>• Data is not local</td>
</tr>
<tr>
<td></td>
<td>• Backup/Recovery is covered by service</td>
<td>• Monthly service fee</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Server</td>
<td>• Locally run by trusted professionals</td>
<td>• IT department or staff needed to support/maintain servers</td>
</tr>
<tr>
<td></td>
<td>• Backup/Recovery is responsibility of IT</td>
<td>• Equipment is expensive to purchase/maintain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Storage</td>
<td>• Data is very accessible</td>
<td>• Disks can be corrupted</td>
</tr>
<tr>
<td></td>
<td>• Data stored locally</td>
<td>• Backup takes time</td>
</tr>
</tbody>
</table>
Data Storage: Procedures

• File Naming
  – Each person has their own filing method
  – Data should be easy to find and use for those that have access
  – Naming convention should be documented and followed by everybody with access to data
Data Storage: Procedures

• Backup/Recovery
  – Set a schedule and stick to it!
    • One missed backup can lead to data loss
  – Plan should be documented
  – Dependent on:
    • Data Storage Location
    • Personnel
    • Equipment
  – Test recovery procedure frequently
SECONDARY DATA ANALYSIS
Using the data

• What data can be used?
  – Consent
  – Approvals
  – Identified vs. De-identified

• How can the data be used?
  – Aggregate data
  – Individual data
What data can be used?

• Potentially any data from completed research studies
  – Aggregate data from reports and publications can be readily used
  – Individual level data needs some additional consideration
Using individual level data

• Any additional use needs to be consistent with the original consent
  – If not, would have to consider re-consent

• Most helpful to have language written for investigators as they are setting up the study
  – Data sharing/ownership plan
  – Consent form suggestions
Using individual level data

• Need to ensure proper approvals
  – Original review boards
  – Tribal research review

• Approvals needed depend on:
  – Whether or not participants can be identified
  – Data sharing/ownership plan
  – Tribal research policy
Using individual level data

• If data is de-identified
  – No longer meets the definition of human subjects research from the Office for Human Research Protections
  – May not need approval from some review boards
  – Would likely need tribal research approval

• If data is identified
  – May need additional approvals for use
  – Would likely need to go back to all original review boards
  – Would need tribal research approval
How can the data be used?

• Need to know what is available
  – Regulatory Knowledge Core of CRCAIH has helped tribal partners catalog previous research
  – Possibility of data repositories

This presentation is supported by the National Institute on Minority Health and Health Disparities of the National Institute of Health under Award Number u54MD008164 (PI – Elliott)
How can the data be used?

• Aggregate data
  – Community assessments
    • Community level data on disease prevalence and risk factors
    • Focus group report on healthcare for children
  – Grant proposals
    • Area specific data
    • Disease prevalence
    • Risk factor prevalence
How can the data be used?

• Individual data (with proper approvals)
  - Changes over time
    • Examine longitudinal trends
    • Examine how tribal priorities fit with research studies
  - Combination of multiple studies
    • Biomonitoring data
    • Behavioral risk factor data
    • Disease prevalence data
  - Combining other data or primary data
Summary

• Data returned for tribal stewardship
  – Accessible
  – Stored securely
  – Preserved
  – Used if possible

• Contact the Methodology Core
  – Questions or advice
  – Assistance
References

• Warren-Mears, V. “Principles and Models for Data Sharing Agreements with American Indian/Alaska Native Communities. (http://genetics.ncai.org/files/Principles%20and%20Models%20for%20Data%20Sharing%20Agreements.pdf)


• http://genetics.ncai.org/sharing-data-and-protecting-your-community.cfm
MC Contact Information

Susan Puumala, PhD
Director
Susan.Puumala@sanfordhealth.org

Chad Birger, MS
Senior Methodology Analyst
Chad.Birger@sanfordhealth.org

Wyatt Pickner, BA, BS
Research Associate
Wyatt.Pickner@sanfordhealth.org

Katie Burgess, MPH
Senior Epidemiology Analyst
Katherine.Burgess@sanfordhealth.org

Kaushal Chaudhary, MS
Methodology Analyst
Kaushal.Chaudhary@sanfordhealth.org