Connecting the Dots:
A Data Sharing Framework for the Local Public Health System

PUBLIC HEALTH INFRASTRUCTURE & SYSTEMS
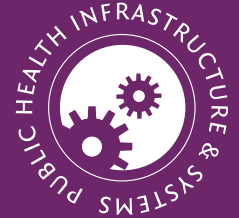
Introduction

Data sharing is defined as "the voluntary provision of information from one individual or institution to another for purposes of legitimate scientific research."[1] While many data sources provide nationally representative data about types, causes, incidence, prevalence, and trends, local health departments (LHDs) often need localized data to measure risks, trends, and the outcomes and impact of local programs, interventions, and laws.[2] Data sharing between LHDs and other local organizations, schools, hospitals, police departments, and judicial systems presents an opportunity for conducting public health research, generating knowledge and exchange, addressing public health issues, and evaluating the success of public health interventions.

The purpose of this document is to translate the Colorado Clinical and Translational Sciences Institute's Data Sharing Framework for use by LHDs. This issue brief provides LHDs, and their partners, valuable lessons learned, best practices, and characteristics of successful data sharing. LHDs, and their partners, should consider the guidelines outlined in this report prior to engaging in collaborations that may involve data sharing activities.

Benefits of Data Sharing

Data sharing provides an opportunity for independent re-analysis and meta-analyses.[3] The Inter-University Consortium for Political and Social Research outlines several other benefits to data sharing:[4]

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinions;
- Promotes new research and allows for the testing of new or alternative methods;
- Improves methods of data collection and measurement through the scrutiny of others;
- Reduces costs by avoiding duplicate data collection efforts; and
- Provides an important resource for training in research.

Despite its benefits, data sharing raises social, ethical, and administrative challenges, and the legal issues surrounding data sharing may present an obstacle, particularly for those who seek access to health information. For instance, the Health Insurance Portability and Accountability Act (HIPAA), which protects the privacy of individually identifiable health information, Food and Drug Administration regulations, and the Federal Policy for the Protection of Human Subjects do not agree on aspects of data sharing.[3] Furthermore, rules on data sharing may vary by state, particularly for disease-specific data.[3] Reluctance or refusal to share data may also be fueled by the perceived "cost and trouble of putting datasets together."[3]

Data Sharing Framework

Developed by the Colorado Clinical and Translational Sciences Institute's Community Engagement Core's Community Health Data and Monitoring Committee, the Data Sharing Framework conceptualizes data sharing within the context of a data life cycle, data management plan, and a data sharing agreement. Figure 1 depicts the relationship between each element.[5] The structure provided by a data management plan and data sharing agreement can enable LHDs to better engage community partners, track trends, measure indicators of a public health problem, or evaluate the impact of a program.
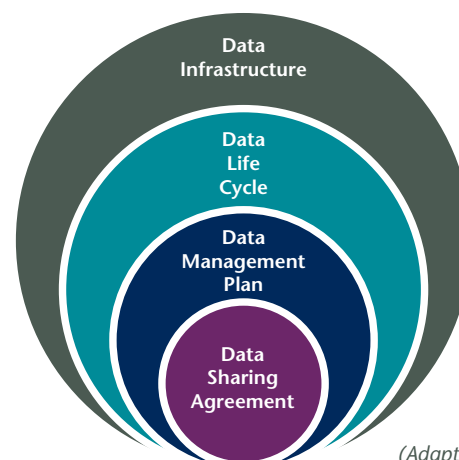
FIGURE 1: DATA SHARING FRAMEWORK

Data Infrastructure · Data Life Cycle · Data Management Plan · Data Sharing Agreement

(Adapted from Jarquín, 2012)

Data Infrastructure

Data infrastructure supports data access, use, and sustainability. Before initiating a project involving data sharing, LHDs should ensure that they use the most up-to-date standards for data sharing and have appropriate staffing, time, and computer systems and software in place to support the project.[6] Community and leadership support are also key components that will facilitate a successful data life cycle, data management plan, and data sharing agreement. Finally, strong partnerships among LHDs and their partners are vital in ensuring that data are collected, analyzed, and disseminated appropriately.

Data Life Cycle

Before developing a data management plan, LHDs must understand the data life cycle. Throughout a project, data go through six phases (Table 1).[4]

Data Management Plan

When possible, LHDs should plan ahead to ensure that plans for data management and data sharing are established in the early phases of a project. LHDs should develop a plan to address data sharing and archival considerations that affect all stages of the data life cycle. A data management plan identifies data to be shared and guidelines for how to process, format, access, and archive data.[7] See Table 2 for recommended elements of data management plans.[4]

Data Sharing Agreement

Creating a data sharing agreement helps to build successful partnerships and supplements an overall data management plan.[5,8] A data sharing agreement should build trust and foster good communication, long-term planning, and open and safe sharing of data.[5,9] The goal of a data sharing agreement is to determine cultural and ethical ways to disseminate data, meet funding requirements, and ultimately improve health outcomes and add to the body of scientific knowledge.[5]

Data sharing agreements should include the following three elements: strong partnership with clear communication, clear process, and well-developed content.[6,9] Clear communication among partners is key for successful data management plans and data sharing. Partners must fully comprehend their role in the partnership, and the needs of LHDs must be balanced with the needs of partner organizations.[9]

A clear process must be outlined to develop and maintain a data sharing agreement. Creating and adhering to the process will benefit the outcomes of the research process.[5] Any funding requirements for data management plans or data sharing agreements must be adhered to and should be built into the data sharing process.

Characteristics of Strong Partnerships

- Serve a specific purpose and may take on new goals over time;
- Have an agreed-upon mission, values, goals, measurable outcomes, and accountability;
- Are characterized by mutual trust, respect, genuineness, and commitment;
- Build upon identified strengths and assets but also work to address needs and increase capacity of all partners;
- Balance power among partners and enable resources among partners to be shared;
- Make clear and open communication an ongoing priority by striving to understand each other's needs and self-interests and develop a common language;
- Have principles and processes that are established with input and agreement of all partners;
- Include feedback among all stakeholders in the partnership with the goal of continuously improving the partnership and its outcomes;
- Share benefits of the partnership's accomplishments; and
- Include a process for closure.

(Adapted from Community-Campus Partnership for Health, 2010)

Finally, data sharing agreements should have well-developed content and should supplement a data management plan. Data sharing agreements typically have content areas that describe the data to be collected, how and by whom the data will be managed, and how the data will be disseminated throughout the lifecycle of the project. While not every content area of a data sharing agreement must be included, Table 3 provides content areas that may be included, considerations, and examples.[5]

TABLE 1: DATA LIFE CYCLE

Phases	Description
Phase 1: Proposal Development and Data Management Plans	Discuss plans for data during proposal development or while the researcher is outlining and writing the grant application. This can simplify the creation of a data management plan and avert problems that may arise. See Table 2 for all components of a data management plan.
Phase 2: Project Start-Up	Once the project has begun, continue to plan for the management of a data collection. Consider the file structure, variables, data integrity, and documentation. Determine what software will be used and how data will be stored.
Phase 3: Data Collection and File Creation	Adhere to best practices for collecting and documenting components of data collection. Clearly document data and check for consistency.
Phase 4: Data Analysis	Carefully clean the data to discover errors and maintain explicit versions of a dataset. For instance, one version may result from the data collection process, while the other may result from data cleaning. Back up datasets frequently to prevent lost work.
Phase 5: Preparing Data for Sharing	Alter files so that datasets respect respondent confidentiality but ensure that alterations to the dataset do not unnecessarily reduce another researcher's ability to analyze the data.
Phase 6: Depositing Data	Organize data in a way that is accessible using a standard statistical package. The decision to archive secondary data should be determined with the original data producer.

TABLE 2: ELEMENTS OF A DATA MANAGEMENT PLAN

Element	Description
Data Description	Briefly describe the information to be gathered and the nature, scope, and scale of the data to be produced.
Existing Datasets	Thoroughly review existing datasets to make clear the value of the proposed research and why currently available datasets are inadequate to answer research questions.
Format	Choose formats of data that are preferred for archiving to make processing and releasing data more efficient.
Metadata	Describe the metadata (e.g., means of creation, purpose, time and date of creation, author) to be provided and discuss the metadata standards used.
Storage and Backup	Describe how and where copies of research files will be stored to ensure their safety.
Security	Describe measures that will be taken to ensure data are secure.
Responsibility	State who will be responsible for the data throughout the data life cycle.
Intellectual Property Rights and Data Ownership	Indicate who will hold intellectual property rights to the data and other information created or collected by the project and how to obtain permission to use or disseminate data.
Access and Sharing	Indicate how data will be archived and shared and why that option has been chosen.
Audience	List the potential secondary users of the data.
Selection and Retention Period	Describe how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.
Archiving and Preservation	Outline the procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data.
Ethics and Privacy	Indicate how informed consent will be handled with respect to informing respondents that the personal information they provide will remain confidential when data are shared or made available for secondary analysis.
Informed Consent	Include the communication processes in place that will allow individuals to make informed choices about participating in a research study.

(Adapted from Inter-University Consortium for Political and Social Research, 2009)

Steps for Creating and Adhering to a Clear Process for Developing and Maintaining a Data Sharing Agreement

- Develop a data sharing agreement early in the project;
- Use the data sharing agreement to develop shared vocabulary to use throughout the partnership;
- Formalize the agreement by writing it down and signing it;
- Meet regularly and revisit the data sharing agreement often;
- Include flexibility with the agreement and make changes if necessary; and
- Identify clear termination dates within the agreement (e.g., termination of data collection, analysis, and dissemination).

(Adapted from Jarquín, 2012)

TABLE 3: CONTENT AREAS OF DATA SHARING AGREEMENTS

Content Areas	Considerations and Examples
Title	Include a brief, descriptive title (e.g., Data Sharing Agreement, Memorandum of Understanding)
Data Overview	Summarize data to be gathered, used, or made available
Resources	Describe funding requirements, infrastructure, and other resources
Intellectual Property	Describe institutional review board, intellectual property rights, ethics, and other legal requirements
Ownership	Describe who owns the data (e.g., partners, local communities, researchers)
Storage	Describe how the data will be securely stored and where the data will be housed
Access	Describe how data will be made available to stakeholders and how confidentiality will be maintained
Approval	Determine approval for which stakeholders have what levels of access to data
Roles	Outline which stakeholders have what roles and are responsible for data collection, storage, management, analysis, dissemination, access, and ownership
Training	Identify training needs and requirements for all partners
Analysis	Determine who will analyze the data and how
Dissemination	Determine all mechanisms for dissemination
Timeline	Determine how long data will be archived, where it will be archived, when data collection will end, when the data sharing agreement will be over, and when and how the data will be disseminated

(Adapted from Jarquín, 2012)

Background

In 2008, Colorado passed the Public Health Act (SB 08-194), which mandated LHDs to complete a community health assessment and community health improvement planning (CHA/CHIP) process every five years. San Juan Basin Health Department (SJBHD) identified obesity as a health priority and needed to determine how to measure accurately the prevalence of childhood obesity in southwest Colorado. SJBHD's vision was to improve child health in the community.

Challenge

Although San Juan Basin is included in surveys conducted by the Colorado Department of Public Health and Environment (CDPHE), the sample size of San Juan Basin residents is often small, yielding uncertainty regarding the prevalence of childhood obesity in the community. Data available to SJBHD included the Colorado Child Health Survey and Healthy Kids Colorado, a written survey administered to children aged 15–18 in Colorado. Molly Gutilla, a DrPH student in Epidemiology at Colorado School of Public Health at University of Colorado, was tasked with developing a systematic and sustainable way to collect data on childhood obesity in the community, using local resources, allowing SJBHD to measure more accurately the prevalence of childhood obesity in the community.

Solution

In 2013, Gutilla completed her practicum at SJBHD and was instrumental in successfully establishing a data sharing plan with a local pediatrics office. Gutilla and SJBHD were interested in collecting data for height, weight, age, gender, insurance status, race/ethnicity, and county of residence of children aged 0–18 years old in San Juan Basin. To obtain these data, SJBHD partnered with a local pediatrics office and SJBHD's legal counsel. SJBHD's legal counsel drafted language for the data sharing agreement, which contained legal language and guidelines on collecting de-identified data, extracting data, and storing data. Gutilla drafted the data management plan, which was more descriptive, detailed, and free of jargon. The data management plan established shared language to help partners communicate. The pediatric office reviewed the data sharing agreement and had the opportunity to ask questions and discuss the agreement.

Facilitators to Data Sharing

Many factors facilitated the process of sharing data. The local pediatrics office served and had access to data from the majority of the children in San Juan Basin. For that reason, SJBHD needed to partner with just one organization to collect the desired data. Moreover, an existing professional and collaborative working relationship was already established between the LHD and the pediatrics office, and the pediatrics office was willing to help SJBHD reach its goal of collecting accurate data on childhood obesity in the community. Furthermore, the owner of the pediatrics office read a journal article on mapping obesity and thus understood and supported SJBHD's project and the need for sharing data. SJBHD received examples of data sharing agreements from CDPHE, and SJBHD's legal counsel helped develop the data sharing agreement.

Challenges to Data Sharing

Challenges faced included writing legal documentation to allow data sharing. Although the state health department willingly provided examples of its data sharing agreements, the documents needed to be adjusted to the local context. It was difficult to find examples of legal documentation in data sharing at the local level. The pediatrics office used electronic health records (EHRs) and Gutilla faced challenges with extracting data, writing queries, and creating data reports. Gutilla reached out to the software technicians of the EHRs, but they were unfamiliar with providing the data in a way conducive to data analysis (i.e., a table with one observation per row with variables in columns). After a lot of troubleshooting, time, and effort, Gutilla eventually was able to get the data in the format she needed.

Outcomes

As a result of this collaboration, SJBHD was able to describe more accurately prevalence of childhood obesity in the community. With baseline data, the health department was better equipped to measure outcomes of its CHIP. The overall description and goals of the project, the plan for dissemination, and ownership of the data were key components of the data sharing agreement.

Lessons Learned

Key elements to success are legal counsel familiar with data sharing agreements, resources and staff to support the process, established working relationships between the LHD and agencies that have data, and EHRs that are created with epidemiologists and public health professionals in mind. While not all of these were in place in San Juan Basin, Gutilla's availability and tenacity and the existence of the working relationship between SJBHD and the pediatric office facilitated the project.
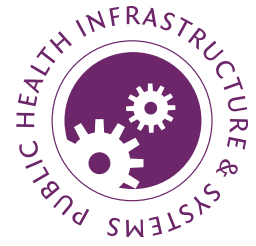
Conclusion

Successful data sharing is supported by a common goal between two or more organizations. The case study illustrates that experts, such as legal counsel familiar with drafting data sharing agreements; examples of data sharing agreements, provided by the state health department or other LHDs; and staff and resources to support the process provide the necessary infrastructure for initiating and maintaining a data sharing agreement and following a data management plan. As indicated in the report, open collaboration; trust between the LHD and its partners; flexibility; tracking of evolving standards for data sharing; and acknowledgment of internal and external priorities of both the LHD and its partners also drive the success of a data sharing agreement.

Shared goals help to clarify partner roles in the broader public health system and can initiate and invigorate partnerships. Moreover, complete understanding of the data lifecycle and the creation and adherence to a comprehensive, well-written data management plan and data sharing agreement provide structure, clear roles and responsibilities, and a defined process that facilitates a successful data sharing project and ethical, responsible use of data.

References

1. Hogue, C.J. (1991). Ethical issues in sharing epidemiological data. *J Clin Epidemiol 44*(Supple 1): 103S–107S.

2. Public Health Informatics Institute. (2005). *The big picture: Developing an enterprise view of public health information systems.* Retrieved March 26, 2013, from http://www.phii.org/sites/default/files/resource/pdfs/bigpicture.pdf

3. National Academy of Sciences. (2013). *Sharing clinical research data: Workshop summary.* Retrieved Aug. 8, 2013, from http://www.nap.edu/catalog.php?record_id=18267

4. Inter-University Consortium for Political and Social Research. (2009). *Guide to social science data preparation and archiving: best practice throughout the data lifecycle (4th Edition).* Retrieved Aug. 27, 2013, from http://www.icpsr.umich.edu/files/icpsr/access/dataprep.pdf

5. Jarquín, P.B. (2012). *Data sharing: Creating agreements in support of community-academic partnerships.* Retrieved Aug. 27, 2013, from http://cctsi.ucdenver.edu/communityengagement/resources/datasharingguidelines/documents/datasharingcreatingagreements.pdf

6. Public Health Informatics Institute. (2010). *Leveraging immunization data in the e-health era: Exploring the value, tradeoffs, and future directions of immunization data exchange.* Retrieved March 25, 2013, from http://www.phii.org/sites/default/files/resource/pdfs/ehr-hie-iis%20report%20-%20oct%202010.pdf

7. Public Health Informatics Institute. (2012). *Redesigning public health surveillance in an eHealth world.* Retrieved Sept. 5, 2013, from http://www.phii.org/sites/default/files/resource/pdfs/requirements%20lab_final%20deliverables_rwj%20surveillance.pdf.

8. Singh, J.A. & Daar, A.S. (2009). Intra-consortium data sharing in multi-national, multi-institutional genomic studies: Gaps and guidance [Perspective]. *The HUGO Journal, 3*, 11–14. doi:10.1007/s11568-009-9130-5.

9. Community-Campus Partnerships for Health (CCPH). (2010). *Principles of good community-campus partnerships* (Adopted by CCPH board of directors, October 1998). Retrieved July 22, 2013, from http://depts.washington.edu/ccph/principles.html#principles

Acknowledgments

FOR MORE INFORMATION, PLEASE CONTACT:

NACCHO's Injury and Violence Prevention Program: injuryprevention@naccho.org

Share your data sharing story at nacchostories.org.