

Brief Meeting Notes

Webinar for Young Environmental Health Investigators on Data Sharing

April 3, 2014

Introduction

Dr. Ana Maria Mora (ISEE Student and New Researchers Network Steering Committee member)

Purpose of webinar was to:

- Provide information, experiences and discussion focusing on the challenges and concerns of young investigators interested in sharing or utilizing environmental health data.

Meeting Overview:

- Summary of Environmental Health Science Data Sharing Workshop held at NIEHS in April 2013: Dr. Kim McAllister (NIEHS).
- Data sharing strategies: Francine Laden (Harvard School of Public Health), Dr. Perry Hystad (Oregon State University) and Dr. Heather Volk (University of Southern California).
- Ethics of informed consent: Professor Wael Al-Delaimy (University of California San Diego, Chair of ISEE Ethics and Philosophy Committee).
- Panel discussion moderated by Dr. Cindy Lawler (NIEHS).

Summary of Previous Environmental Health Science Data Sharing Workshop and Request for Information (RFI)

Kim McAllister (NIEHS)

- The **overall goals** of RFI and workshop were to explore the unique challenges associated with environmental health data sharing, clarify the needs and goals for data sharing, and identify the best practices and approaches of data sharing applicable to environmental health.
- **Unique considerations** when sharing environmental data include: heterogeneity of measurements (i.e., exposure measurements), potential to identify people from geographical data, and increased interest in return of individual or community-level research results, regulatory implications, concerns of vulnerable populations.

- **Importance of Community Involvement:** Importance of CBPR, consultation and informed consent.
- **Institutional Review Boards:** lack of consistency and clarity.
- **Legal and Regulatory Concerns:** re-analysis or re-interpretation of data in effort to delay reform or influence litigation or public opinion.
- **Protection of privacy and confidentiality:** Concerns regarding the identification of sources of exposure and potential discrimination against communities. Solutions to this problem involve “anonymizing” exposure data using unique subject identifiers and/or archiving data in separate external repository.
- **Computational challenges:** Pace of data generation is surpassing infrastructure for storage, analysis and dissemination.
- **New NIH Genomics Data Sharing Policy:** a nice model for those using environmental science data. Brings together human and non-human genomic data under single policy. The goal will be the broad sharing of data.
- **Data sharing falls under Goal 7 of the NIEHS Strategic Plan--**NIEHS has focused on:
 - Developing common environmental measurement vocabularies as evidenced by the June 2013 NIEHS workshop and the upcoming webinar.
 - Promoting various databases such as the Comparative Toxicogenomics Database (CTD).
 - Developing guidelines for data sharing plans.
- **Seeking perspectives of younger investigators** who were not as well represented at the workshop and RFI.

Data Sharing Strategies in Environmental Epidemiology

Senior Researcher Perspective: Dr. Francine Laden (Harvard School of Public Health)

- Shared her extensive experience sharing data from Nurses’ Health Study and other consortia efforts focused on environmental epidemiology of cancer, cardiovascular and respiratory disease with a focus on air pollution.
- Recommended looking at the data sharing plans for large cohort studies to get a better understanding of how they have handled data-sharing.
- Why share data?—maximizes the scientific value of epidemiological data.

Steps for requesting access to large cohort data

- Step 1: Review study websites and publications, and make sure no overlap of your proposal with study you want to work with.
- Step 2: Find affiliated sponsor to collaborate with and review proposal with them.

- Step 3: Submit proposal to be reviewed.
- Step 4: If approved, apply for funding.
- Environmental data has extra challenges because location-based data is identifiable data. Personal identifiers need special attention!

New Investigator Perspective:

Dr. Perry Hystad (Oregon State University)

- Shared his experiences as both a student and a new researcher establishing collaborations with several national and international health projects (e.g. the NECSS, CHILD and PURE studies).
- Student perspective
 - Used NECSS case-control study.
 - Development of new collaboration from scratch took time (2 ½ years from first contact with PI to obtaining full data set!)
 - Required a personal relationship with PI: phone calls and in-person visits.
 - There are ample opportunities to create new research and knowledge with existing data!
- (Very) New investigator perspective:
 - Data access is number 1 need and biggest hurdle for new researchers.
 - Current Research Collaborations: 1) PURE-AIR Cohort (air pollution and cardiopulmonary disease in existing PURE cohort) and 2) CHILD study (prenatal air pollution exposure and fetal and infant growth).
 - Advantages of large study: data-sharing, publication and collaboration plans already defined.
 - Disadvantages of large study: tight budget.
 - Lessons Learned:
 - Use established networks/relationships (i.e., your PhD advisor) to contact study PI early on.
 - Build relationships and trust - visit the study site in person!
 - Focus on collaborations and your contributions rather than just data access.
 - Be persistent.
 - Reciprocate (share your own data!)
 - Challenges: tight budgets; lack of incentive to share data and collaborate; most studies do not address environmental aspect.
 - Opportunities: new investigators should promote data sharing culture; metric of investigator success should include sharing data.

Dr. Heather Volk (University of Southern California)

- Shared experiences working with shared data from individual investigators to repository datasets.
- Uses data from the CHARGE Study to look at air pollutants and Autism Spectrum Disorders.
- Experience building a collaboration:
 - Have specific, clear goals of what your collaboration will bring to the study.
 - Send a written proposal.
 - Phone calls and visits important to PIs.
 - Develop secure data transfer plans.
 - Keep in mind data-sharing requirements and the confidentiality and ownership issues of the data you are generating.
- Experience with a biorepository:
 - Worked with AGRE (Autism Genetics Resource Exchange) repository/database which contained 1,736 pedigrees, DNA and serum biosamples, and a phenotyping battery.
 - No environmental data were associated with biosamples so had to find out where children were born and where mother lived during pregnancy.
 - Involved both surveying families and a computer search for past residences, which required additional data sharing negotiations. Working with a commonly used biorepository makes continuing collaboration & data sharing easier.
- Future Goals/Challenges:
 - How to use large databases or repositories when they are not created for environmental research?
 - Need to consider data ownership when collaborating: Who to share with and how much to share?

Ethical Perspectives of Data Sharing

Professor Wael Al-Delaimy (University of California, San Diego)

- Ethical challenges:
 - Inconsistencies in Institutional Review Boards (IRBs) across states and countries.
 - Conflicts of interest are an obstacle of wide-spread data sharing.
 - Overcoming ownership (i.e., publication credit) issues.
 - Lack of support for open consent.
 - Lack of infrastructure to handle and/or secure data.

- Limited biosample availability (how are banked samples prioritized for sharing?)
- Risk of data sharing:
 - Loss of privacy.
 - Infringement of subject autonomy.
 - Inappropriate study design (using primary study design for secondary research questions).
 - Doubt manufacturing (e.g., by industry).
 - Loss of proprietary data privileges (e.g., industry loses opportunity for profit from data).
- Benefits of data sharing:
 - Health advancement and prevention.
 - For researchers in developing and lower-income countries.
 - Cost effectiveness.
- Different data types have special issues:
 - Questionnaire/written data: Re-analyze original data versus new hypothesis.
 - Biomonitoring/bank data: Exposure assessment (new findings or re-analyzing existing); Genomic/omic assessment (new findings or, re-analyzing existing).
 - Linking datasets to create new dataset: e.g., Gene x Environment or medical records can create confidentiality/privacy issues.
- Stakeholders
 - Researchers:
 - Original PI has ownership of data but also the burden and cost of data sharing.
 - Secondary PI needs to determine if original study design can address new question, determine if there is a conflict of interest, how to share costs. Need to ask for consent of original subjects.
 - Subjects/communities:
 - Subjects can volunteer their data but do not own it. They do have autonomy to decide which study uses data and can ask to see results and feedback.
 - Subjects may not have consented if they were asked for open consent for future studies.
 - Results and feedback are expected.
 - Involvement of community is required if community can be identified (e.g. Havasupai Indians), but this is not always the case.
 - Communities can suffer stigma and environmental injustice.
 - Time consuming process; can delay data sharing.

- Data-sharing organizations & associated challenges:
 - HMO's & clinical entities – complex HPPA rules.
 - Registries, Departments of Health at the state and county level – inefficiencies, lack of resources & staffing.
 - Federal agencies such as the EPA and the CDC – political challenges.
 - Universities – intellectual property issues.
 - Industry – conflict of interest and intellectual property issues.
- Data Sharing Access:
 - Data Users: Environmental sciences need it most.
 - Data Protectionists: Concerned with the increase in identity theft and selling of personal information.
 - Data Obstructionists: Industries (e.g. cell phone companies).
- International Setting:
 - Challenges: Regulation differences, lack of local IRBs and trained scientists, transfer of data across borders.
 - Human genome is an example of successful international data sharing.
- Neutra et al 1996 Re-analyses Guidelines
- Conclusions:
 - Data sharing is a necessity - Not “if” but “how” data should be shared (Institute of Medicine, 2013).
 - The benefits of data sharing outweigh the risks.
 - Need for guidelines, ongoing public engagement (funding agencies like NIEHS are an appropriate platform for this).

Panelist Question and Answer Session

Moderated by Kim McAllister and Cindy Lawler

1. When trying to set up a new data sharing collaboration, is it generally assumed that you will also contribute data? If not, what do you usually need to contribute (other than your time)?

Laden: *Be focused on and specific about what you can bring to the study (e.g., a novel research question and your willingness to work including time, motivation, and expertise such as spatial exposure assessment); agreed with Hystad and Volk about building relationships and importance of collaboration.*

2. What kinds of funding did the panelists apply for when using shared data from a study?

Hystad: *Apply for early career/new researcher awards like the Early Independence and Outstanding New Investigator awards or pilot funding.*

Volk: *Applied for almost every grant available including pilot projects at University, private foundations such as Autism Speaks, Health Effects Institute, R21, R01. She liked Perry's idea of applying for early career awards and grants.*

3. How do you go about negotiating authorship issues when you enter into these large collaborative projects? What about authorship issues with respect to those who initially generated data versus those who come in with a new idea for re-analysis?

Laden: *Discuss authorship early-on! When starting work on a new project, decide on a first, second, and last author upfront. First author typically comes up with the idea and does the work; the last author oversees work and provides funding; can also be a senior investigator or first author mentor. For pooling projects, working groups may decide authorship.*

Al-Delaimy: *Authorship is an ethical issue dictated by the original data owner. Deal with the issue early on!*

4. Two questions from a student: How do you incorporate publically available data into a project? How do you communicate results with the public?

Volk: *Consider the quality of publically available data. Regarding communication, she was an Environmental Health Network science communication fellow for one year which helped her learn how to communicate with reporters and to the public. She also suggested working with Core Center COECs.*

Al-Delaimy: *If information is controversial, work with the data owners.*

5. If the data cannot be directly shared (i.e. transferred), what are other possible strategies for collaborating (e.g., all partners run local analyses and the results are subsequently meta-analyzed centrally, one (the lead) researcher travels from one partner to the next to run the analyses, etc...)?

Hystad: *If data can't be transferred, new technology is available such as secure VPN or secure roaming server. This seems to be becoming more common.*

Laden: *People from outside can access data on the data sharer-side through the VPN.*

6. What are some models for Community-Based Participatory Research (CBPR) Projects?

McAllister: *PEPH program at NIEHS.*

Al-Delaimy: *CBPR is at the heart of environmental epidemiology. Can be a lengthy process. Make sure you are not in controversy so that you maintain your independence as a researcher. Get community involved in issue from the beginning and in the hypothesis generation and data sharing conversations.*

7. Question for Drs. Hystad and Volk: As a new researcher, how do you think about the balance/transition of focusing on existing data versus the creation of new data, especially in the context of trajectory toward tenure?

Volk: *Has had most success with using existing data sources; funding is tough. Securing funding for novel data collection requires building a convincing track record first.*

Hystad: *Primary data collection for large scale studies is hard to start with; agree that funding for novel data collection is difficult.*

8. How can new researchers create a data sharing culture in environmental health?

Laden: *Make sure new researchers know about data-sharing issues at the very beginning.*

Hystad: *New researchers need to understand that they will need to share data and to not fear sharing data. Need a culture shift to enhance data sharing.*

Al-Delaimy: *Data sharing is migrating to the environmental sciences. Researchers using shared data need to understand perspective of original data owners and give them credit.*

Volk: *Tell junior investigators to work together (e.g., work with your friends).*