

Data and Biospecimen Sharing: Setting the Stage

October 16, 2015

Susan M. Pinney, PhD

Department of Environmental Health

University of Cincinnati



New Online Resources! Human Biomonitoring: Principles and Best Practices

*Created by the
Integrated Health Sciences
Facilities (IHSF) Core of the
University of Cincinnati
Center for Environmental Genetics,
with funding from the National Institute of
Environmental Health Sciences
P30 ES006096*

- Video #1 (30 minutes): Biomonitoring Study Design Principles
Susan M. Pinney, PhD
Department of Environmental Health / Center for Environmental Genetics
University of Cincinnati College of Medicine
- Video # 2 (35 minutes): Archiving Biospecimens (Things to Think About)
Jeanette Buckholz, RN, MSN
Department of Environmental Health / Center for Environmental Genetics
University of Cincinnati College of Medicine
- Recommended Readings
- Fernald Community Cohort Biospecimen Processing Protocol (Sample)
- Material Transfer Form Template (Sample log)
- Freezer Assessment at Time of Sample Pull (Sample log)

Available at <http://eh.uc.edu/ceg/human-biomonitoring/>





"It's time I downloaded some information to you."

Here's the situation:

- You have worked for YEARS on a research project.
 - Obtaining the funding
 - Recruiting the population
 - Collecting the data – all kinds of data
 - Collecting and processing the biospecimens
 - Conducting genetic, genomic, proteomic, etc. studies
 - And now are just getting to the point of doing data analyses....
- AND, you are required to SHARE your data and biospecimens!



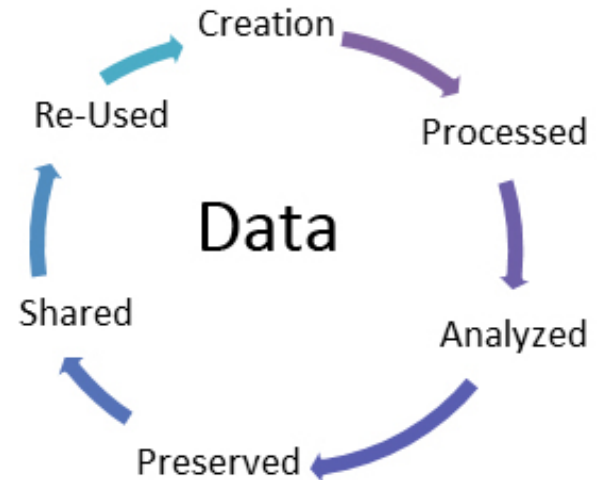
What thoughts do you have?

Responses from audience.

- Frustrating
- Protective of own data
- Do not lose oversight
- Balance both sides: first publish, then sharing; do need to share after publishing paper to benefit others
- Sharing is important, funding from government, but research subjects contributed to information and biospecimen. Protecting research subjects, benefit the community?
- Consent of sharing
- Logistics and cybersecurity of data sharing
- The way of doing science, how you can share thinking of science and research participants
- (De)identified data, but 10 years later, only little or no identification seen by other researchers.
- Acknowledgment of the original researchers.
- Identification kept not known

Why should I share data... For my science to move forward

- More collaborations
 - More scientific friends
 - More new ideas
- More publications
- More recognition
 - More invitations to conferences, etc.
- Better positioned to get more funding



Why should I share data.... For better science

- Re-establishes the culture of open scientific inquiry
- Moves science forward more quickly
- Encourages diversity of analysis and opinion
- Promotes new research, testing of new or alternative hypotheses and methods of analysis - *seeing the data with fresh eyes!*
- Permits the creation of new datasets by combining data from multiple sources
- Facilitates education of new researchers — *getting those first publications!*
- Enables the exploration of topics not envisioned by the initial investigators — *one investigator never can do all of the analyses that could be done.*



Taxpayers have paid for research....



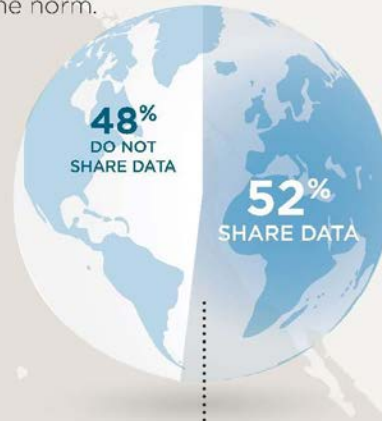
- Maximize the benefit of taxpayer dollars
- Share resources including questionnaires, methods, communications
- Share composite data with study participants

Survey conducted by Wiley







- Researcher view of data sharing
- Contacted 90,000 researchers, wide variety of disciplines
- 2250 responses from those engaged in active research
- <http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/>

GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.



WAYS DATA IS SHARED

-  **67%** As supplementary material in a journal
-  **37%** Personal, institutional or project webpage
-  **26%** Institutional data repository
(i.e. university or institute-sponsored)
-  **19%** Discipline-specific data repository
-  **6%** General-purpose data repository
(e.g. Dryad, figshare)
-  **5%** Other

Globally, researchers also report sharing their data in limited and non-permanent ways: 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).



Where Health Scientists share their work:

- 68% As supplementary material in a journal
- 29% Personal/institutional/lab webpages
- 29% Institutional data repositories
(i.e. *university or institute-sponsored*)
- 21% Discipline-specific data repositories
- 5% General-purpose data repositories
(e.g. *Dryad, figshare*)

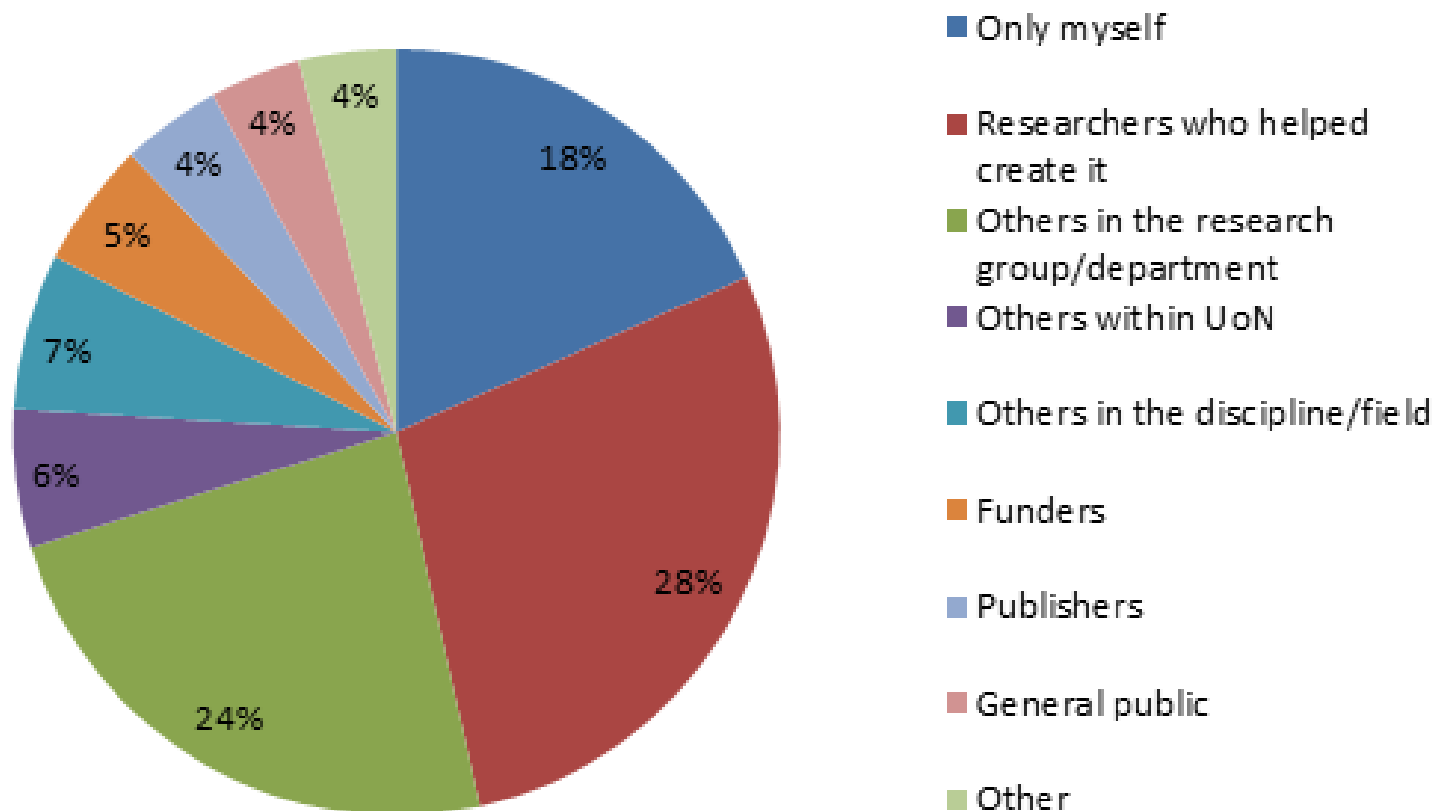
A typical *Health Science researcher* says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assuaged.



REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

- 42%** Intellectual property or confidentiality issues
- 36%** My funder/institution does not require data sharing
- 26%** I am concerned that my research will be scooped
- 26%** I am concerned about misinterpretation or misuse
- 23%** Ethical concerns
- 22%** I am concerned about being given proper citation credit or attribution
- 21%** I did not know where to share my data
- 20%** Insufficient time and/or resources
- 16%** I did not know how to share my data
- 12%** I don't think it is my responsibility
- 12%** I did not consider the data to be relevant
- 11%** Lack of funding
- 7%** Other

People who can access the research data created

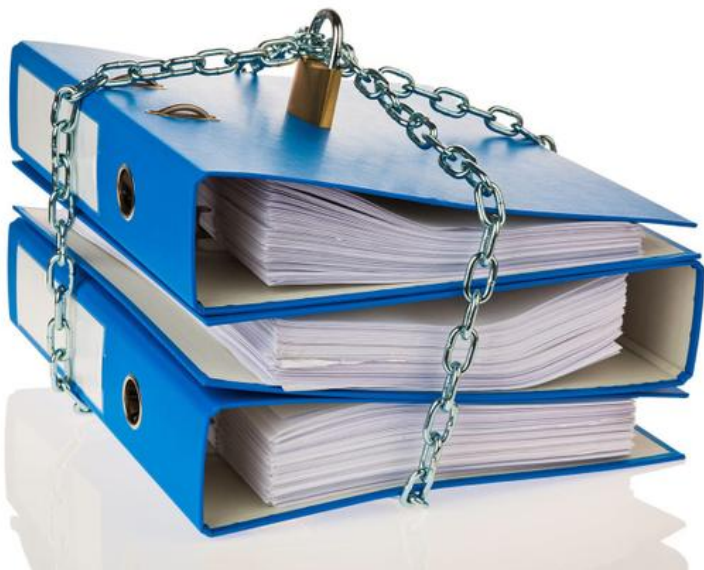


What do funders want?

- Timely release of data
- At time of publication
- Open data sharing
- Minimal or no restrictions if possible
- Preservation of data
- Typically 5-10+ years if of long-term value
- NIH Genomic Data Sharing Policy
 - Became effective January 25, 2015
 - <https://gds.nih.gov/>
 - https://gds.nih.gov/pdf/supplemental_info_GDS_Policy.pdf

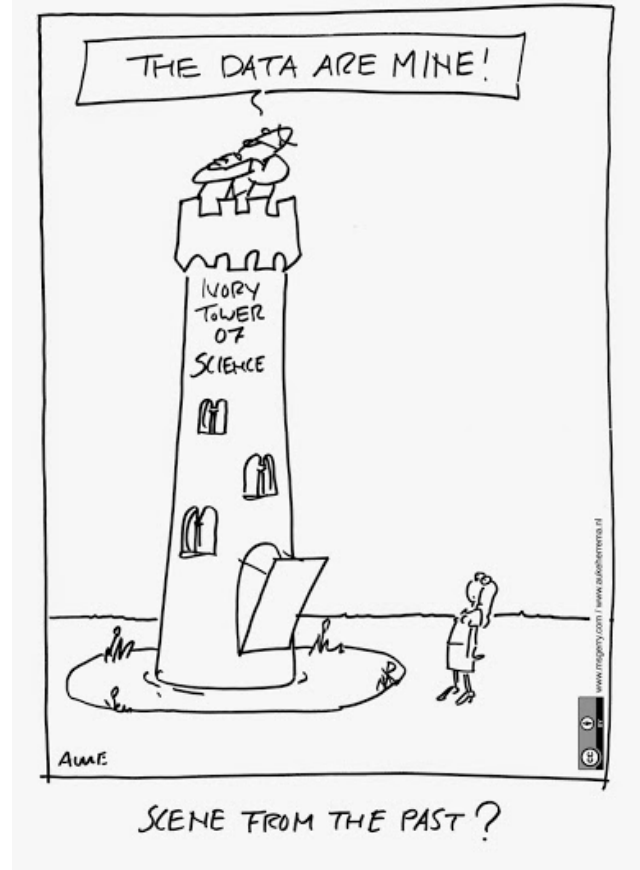


Not this.....



NIH Policy

- Mandatory Genomic Data Sharing for NIH funded research (Scott Langevin)
- Plan (with grant submission) must include:
 - Source of the data and the type of genomic information
 - Data repository – where the data will be submitted and if/how access will be restricted
 - When data will be submitted and released
 - IRB Assurance of the plan (Institutional Certification)
 - Appropriate Data Use – justification for any data sharing restriction)
 - Possible request for exception to submit human genomic data if study will not meet NIH Institutional Certification criteria



“THE DATA ARE MINE” is a thing of the past....

Original Resource Sharing Plan

RESOURCE SHARING PLAN PAR-15-104 Familial Lung Cancer Registry Genetic Epidemiology of Lung Cancer Consortium

University of Cincinnati
Karmanos Cancer Institute
Dartmouth College
Harvard University

Louisiana State University
Mayo Clinic Rochester
Medical College of Ohio
National Human Genome Research Institute

Proposed Data Sharing Plan

The objective of the Familial Lung Cancer Case Registry (FLCCR) within the Genetic Epidemiology of Lung Cancer Consortium is to provide research resources to be used to conduct studies to identify susceptibility genes in familial lung cancer (FLC) that can lead to strategies for the prevention, control, and clinical management of the disease. As required by NIH rules, we will make the data collected as part of this protocol available to outside investigators, but within the limitations of preserving the anonymity of individuals for whom we have genome wide scan data, since these data could theoretically identify a study participant.

In order to maintain compliance with HIPAA regulations, our preferred method will be to execute a data sharing agreement with the requestor for a limited use dataset as defined by the US Department of Health and Human Services (DHHS). We will also encourage, but not require, collaborative use of the data. As with all datasets, there are many subtle nuances to the coding and interpretation of the data that cannot be completely documented in coding books or comment fields. Through a collaborative agreement, we feel that the quality of secondary analyses derived from these data can be greatly enhanced. Alternately, we will provide HIPAA compliant data sets for individuals who are unable or unwilling to execute the data sharing agreement, stripped of identifying information as defined by the US DHHS.

Data will be made available approximately 120 days following the publications of the main findings related to that dataset. Requests for data will be handled on a case-by-case basis and we have allowed 100 hours of analyst time in each year of the funding period to process the requested datasets. For consortium use, our data are posted on a secure website which can be accessed by other GELCC researchers by using individual passwords. We will use this same website to make data available to outside investigators for sharing post publication. Final datasets available to outside investigators will include reported demographic, family structure, cancer affection status (including level of verification) and smoking history data, and laboratory data from tumor specimens. Genotypic data will be made available to outside investigators with agreement to strictly maintain compliance with confidentiality commitments. Outside investigator access to the GELCC secure website will be limited to the dataset prepared for that investigator.

Because this is a longitudinal genetic cohort, we will be collecting and maintaining participant identifying information. A familial lung cancer case cohort with related controls will necessarily include family relationships. Although the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the strong possibility of deductive disclosure of subjects and associated familial relationships with unusual characteristics. Thus, we will follow the data enclave model that make the data available to users only under a data sharing agreement that provides for: (1) a detailed summary of the proposed research project, including a complete list of data requested, (2) a commitment to use the data only for research purposes, (3) a commitment to maintain confidentiality of the data and not to identify any individual participant, and (4) a commitment to secure the data summary statistics using appropriate computer technology, and (5) a commitment to destroy or return to the user data after analyses are completed (6) a commitment to share the findings of statistical analyses with the GELC consortium. The GELC consortium will develop an internal committee to review requests for the use of data according to these stipulations.

All consents used since 2009 include a specific option to allow sharing of genotype and de-identified phenotype with the GWAS data repository. We have re-contacting all previous living study participants to re-consent them with the same option to allow data sharing for sequencing studies. We also will obtain re-consent from family members of deceased lung cancer patients in the case cohort.

Our data sharing policy will be reviewed annually at GELC Consortium meetings, and more frequently if necessary, and necessary revisions made as implications of data sharing policies for genome wide studies become known. Proposed revisions will be presented to NCI, and their approval will be obtained before revisions are implemented.

- Genetic Epidemiology of Lung Cancer Consortium
- Resource Sharing Plan submitted September 11, 2015 with ARA request.
- **Thought I knew what was required but I did not!**
 - Data sharing agreement with requestor
 - Limited use dataset
 - Collaborative agreement
 - Available 120 days after publications of main findings
 - Enclave model because of familial relationships

Revised Resource Sharing Plan

- Email from NCI on 9/28
- Hi Susan,
- Aim 3 will require you to follow the NIH genomic data sharing policy; therefore, we will need a revised data sharing plan. Please use the attached template for the revisions. We cannot forward this for approval until we receive the revised plan.
- Sent revision
- Dear XXXX,
- Attached is our revised Resource Sharing Plan, which now should be in compliance with the NIH genomic data sharing policy. Thanks for bringing this to our attention. Susan
- Email from NCI on 10/6 (as I am leaving for Human Genetics meeting)
- Hi Susan,
The revised data sharing plan can't be approved as it is written.....

RESOURCE SHARING PLAN PAR-15-104 Familial Lung Cancer Case Registry Genetic Epidemiology of Lung Cancer Consortium

University of Cincinnati
Karmaros Cancer
insale
Dartmouth College
Harvard University

Louisiana State University
Mayo Clinic Rochester
Medical College of Ohio
National Human Genome Research Institute

Proposed Resource and Data Sharing Plan

The objective of the Familial Lung Cancer Case Registry (FLCCR) within the Genetic Epidemiology of Lung Cancer Consortium is to provide research resources to be used to conduct studies to identify susceptibility genes in familial lung cancer (FLC) that can lead to strategies for the prevention, control, and clinical management of the disease. As required by NIH rules, we will make the data collected as part of this protocol available to outside investigators, but within the limitations of preserving the anonymity of individuals for whom we have genome-wide scan data, since these data could theoretically identify a study participant. Data produced through this award will be shared through an NIH data repository, consistent with data-sharing under the NIH Genomic Data Sharing Policy (NOT-OD-14-124). The Institutional Certification is underway and will be provided prior to data submission.

In order to maintain compliance with HIPAA regulations, our preferred method

Specifically we propose to make the following information available via controlled access to investigators from scientific institutions that submit Data Access Request (DAR) packages that are reviewed and approved by the NCI Data Access committee.

1. Study documents. We will make available all questionnaires and data collection protocols to other investigators.
2. Summary-level information and aggregate genotype data, including allele frequencies by case-control status, association tests odds ratios, and p values for each SNP in the scan.
3. Individual-level data (genotypes and phenotypes)

Individual-level Genotyping data will include (if applicable):
a. the genotypes calls

Individual-level Sequencing data will include (if applicable):
a. Variant Call Format (VCF) file or equivalent
b. Binary Alignment/Map (BAM) files

Individual-level Phenotype data will include:
a. Case/Control status,
b. Gender,
c. Age in 5 year intervals at the time of diagnosis or designation as control and
d. Family history (defined as like phenotype in first degree relatives) if available

Data will be shared for all of the study participants included in these analyses.

We will share the genotype/sequencing and phenotype data immediately after the genotyping/sequencing data have been cleaned. We expect the cleaning process to be complete 6 months after genotyping/sequencing is finished, which we expect to happen in the 5th month of the grant. All consents used since 2009 include a specific option to allow sharing of genotype and de-identified phenotype with the NIH GWAS data repository. We have re-contacting all previous living study participants to re-consent them with the same option to allow data sharing for both genotyping and sequencing studies. We also will obtain re-consent from family members of deceased lung cancer patients in the case cohort. All consents going forward will specify broad sharing of both genotype and phenotype data with the NIH GWAS data repository.

Formatted: none, Indent Left: 0", Line spacing: single
Formatted: Left: 0.5", Right: 0.5", Top: 0.5", Bottom: 0.5"
Formatted: Not Expanded by / Condensed by
Formatted: Line spacing: single
Formatted: Right: 0"
Formatted: Condensed by: 0.6 pt

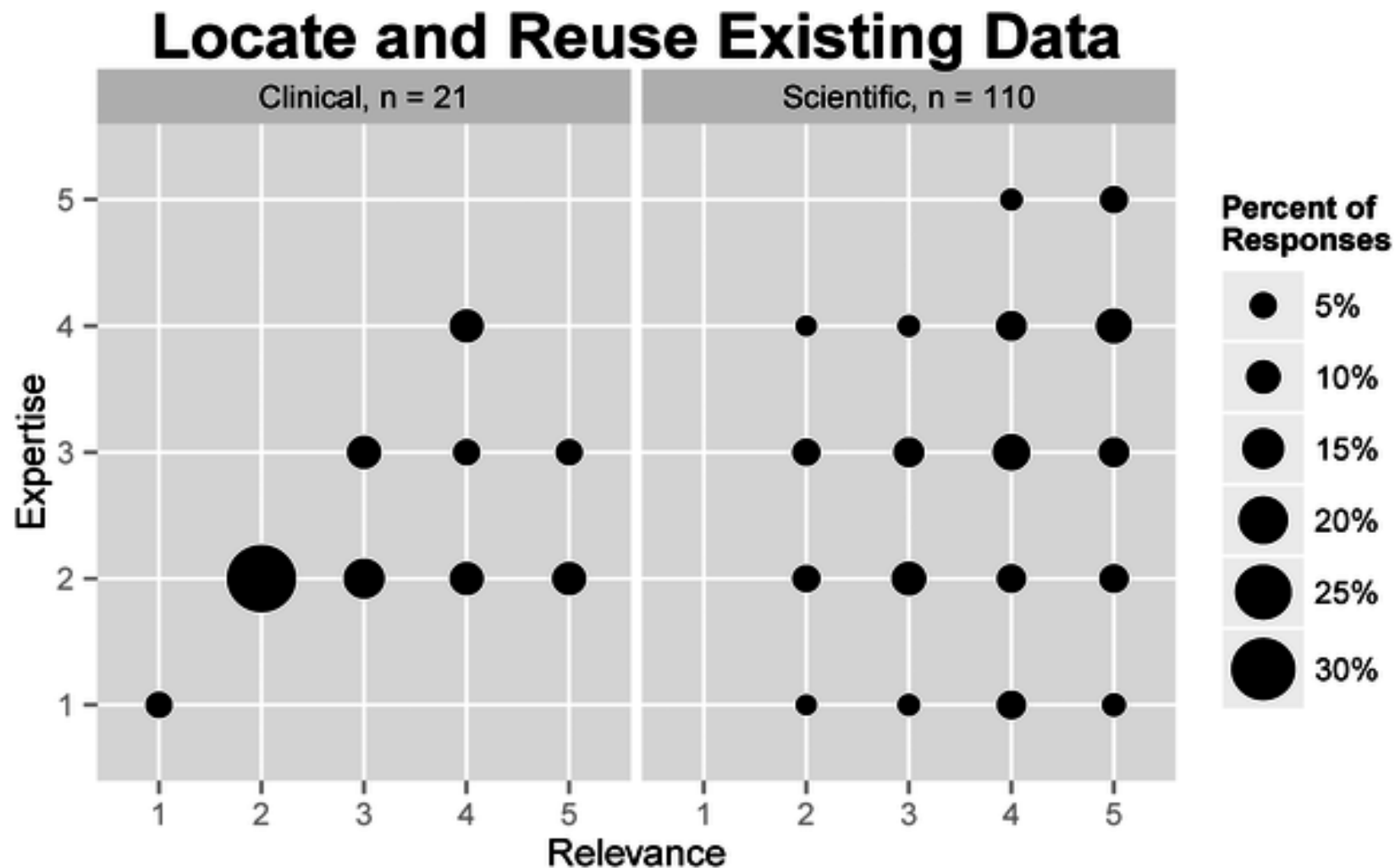
Formatted: Indent Left: 0", Right: 0"

Formatted: Not Expanded by / Condensed by

Federer LM, Lu YL, Joubert DJ, Welsh J, Brandys B (2015) **Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff**. PLoS ONE 10(6): e0129506.
doi:10.1371/journal.pone.0129506

- NIH Intramural Program researchers
- Clinicians and basic science
- Attitudes toward and experiences with sharing and reusing research data
- 135 researchers in analysis
- Rated expertise with data sharing and reuse as low
- Relevance of using shared data also was low
- **Is re-use or sharing data valued in medical research?**

Fig 1. Comparison of self-rated relevance and expertise regarding reusing data among clinical and scientific research staff.

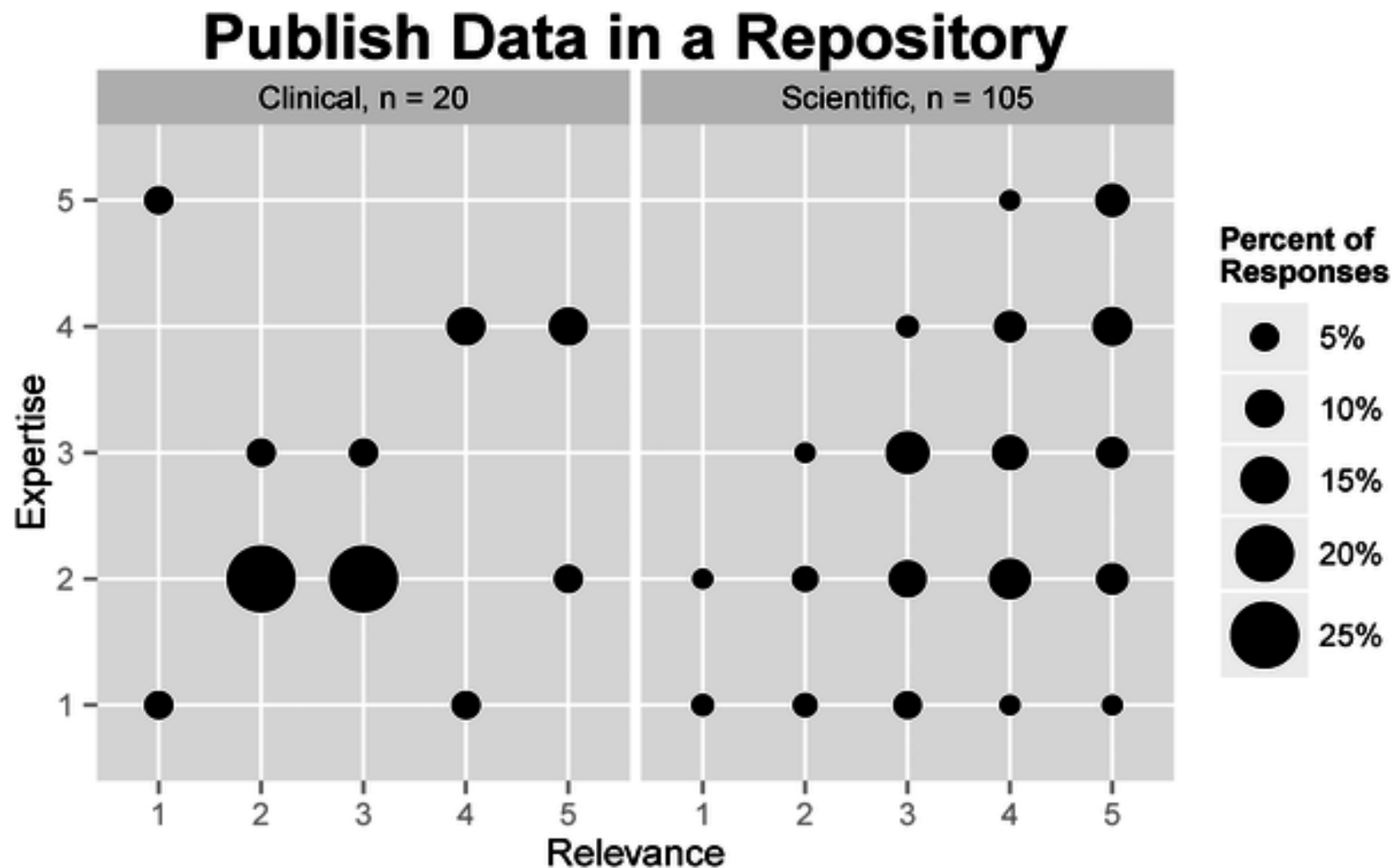


Federer LM, Lu YL, Joubert DJ, Welsh J, Brandys B (2015) Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff.

PLoS ONE 10(6): e0129506. doi:10.1371/journal.pone.0129506

<http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0129506>

Fig 2. Comparison of self-rated relevance and expertise regarding sharing data in a repository among clinical and scientific research staff.



Federer LM, Lu YL, Joubert DJ, Welsh J, Brandys B (2015) Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff.

PLoS ONE 10(6): e0129506. doi:10.1371/journal.pone.0129506

<http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0129506>

To Share or Not to Share? A Survey of Biomedical Researchers in the U.S. Southwest, an Ethnically Diverse Region

Oushy MH, Palacios R, Holden AEC, Ramirez AG, Gallion KJ, O'Connel MA. 2015; PLOS One 10(9); e0138239

- Cancer health disparities research needs access to biospecimens from diverse racial/ethnic populations.
- Investigated barriers, concerns and practices for sharing biospecimens/data among researchers working with minority populations in a 5 state region of the US (Arizona, Colorado, New Mexico, Oklahoma and Texas)
- Emailed survey invitations to 605; 112 responses
- Mostly PIs at PhD granting institutions
- Most were non-Hispanic White (63.4%) and men (60.6%)
- Survey contained questions regarding a virtual biospecimen repository
- Published online on 9/17/15
- Findings:
- Lack of access to sufficient biospecimens
- Limited availability of diverse tissue samples
 - Barriers
 - Poor annotation of biospecimens
- Unwillingness to share
- 50/112 indicated willingness to participate in a virtual repository

Question with 5 point scale, ranging from “very likely = 5” to “very unlikely = 1”.

Table 5

Contrasts comparing different type of researchers against all others in their willingness to share different types of information (scale means).

Scale ^b	Researcher type ^a											
	Basic			Translational			Clinical			Epidemiology		
	Basic	others	<i>p</i>	Trans	others	<i>p</i>	Clin	others	<i>p</i>	Epi	others	<i>p</i>
SI. Specimen information	4.05	2.24	0.01	2.10	3.85	0.01	2.79	4.50	<i>ns</i>	2.88	4.60	<i>ns</i>
SII. Donor information	1.94	2.16	<i>ns</i>	2.29	1.92	<i>ns</i>	2.02	2.00	<i>ns</i>	1.98	2.40	<i>ns</i>
SIII. Grant information	1.40	2.43	0.01	2.15	1.91	<i>ns</i>	2.11	1.33	0.05	2.02	2.50	<i>ns</i>

- **Basic researchers more likely to share specimen information** on a virtual national biorepository compared with non-basic researchers (4.05 vs. 2.24, $p < 0.01$)
- **Translational researchers, clinicians and epidemiologists were less likely to share specimen information** than non-translational researchers (2.10 vs 3.85; $p < 0.01$).
- **Basic researchers and epidemiologists were less likely to share grant information** than non-basic researchers (1.40 vs. 2.43, $p < 0.01$)
- **Clinical researchers were more likely to share grant information** than non-clinical researchers (2.11 vs. 1.33, $p < 0.05$).

Table 8

Researcher requirements for collaborating and sharing data.

Themes	% of 112 total participants	All (46 ^a)	Males (22)	Females (24)	NHW ^b (35)	Minority (10)
Collaboration and acknowledgment	12.5	14 (30.4 ^c)	5 (22.7)	8 (33.3)	12 (34.3)	1 (10.0)
Expertise in tissue research	11.6	13 (28.3)	6 (27.3)	7 (29.2)	11 (31.4)	2 (20.0)
Compliance with institutional and federal policies	8.9	10 (21.7)	5 (22.7)	4 (16.7)	8 (22.9)	1 (10)
Data sharing policies	6.2	7 (15.2)	2 (9.1)	5 (20.8)	5 (14.3)	2 (20.0)
Preservation of resources	4.5	5 (10.9)	2 (9.1)	3 (12.5)	3 (8.6)	2 (20.0)

^ANumber of participants who answered survey questions

^BNHW, non-Hispanic white

^CPercentage of participants who ranked barrier calculated based on demographic category of respondent

Researchers' requirements for collaborating and sharing specimens

The top five themes for requirements reported by the study sample.

Females and non-Hispanic Whites – collaboration and acknowledgment are important.

Males and non-Hispanic Whites – compliance with policies

Females and minorities – Data sharing policies

Minorities – preservation of resources

Why data-sharing policies matter

Guttmacher AI, Nabel EG, Collins FS. 2009; PNAS 106(40).

- Model of investigator owning data increasing replaced by one in which society owns the data
- Broad access accelerates and empowers scientific investigation to benefit society.
- Investigator profits from
 - value added to the data in its deposition in a community database
 - collaborations that the wider data availability attract
- Interests of investigator and of the study participant require protection
- All members of the research community must play an active role in protecting the rights of both research participants and principal investigators.



Data and Biospecimen
Sharing is a
collaboration:

Respect the “rules” of
collaboration.

#1, 6, 7, 8

Collaborations and Your Career^{*}

1. Thou shalt seek individuals with the same passion.
2. Thou shalt get to know your colleagues. They can't read your mind.
3. Thou shalt be a friend to collaborators. Respect their other position responsibilities, deadlines, and non-work related responsibilities.
4. Thou shalt ask colleagues for initial peer review of manuscripts, grants and reports, as a way to begin collaborations.
5. Thou shalt work with friends when possible. They will do things for you that colleagues will not.
6. Thou shalt look for collaborators beyond your usual borders, other colleges within your university, and those at other locations.
7. Thou shalt recognize one's own expertise and respect the expertise of collaborators. Don't assume that you have the expertise of another discipline.
8. Thou shalt clearly define roles, responsibilities, and deadlines. Everyone has an important role and they should know what that is.
9. Thou shalt establish a longitudinal, efficient meeting schedule.
10. Thou shalt acknowledge/cheer accomplishments of collaborators.



*Developed at the Collaborations and Careers Workshop, May 7, 2014, sponsored by the Career Development Core and the Integrative Health Sciences

