

Summary of Request for Information (RFI) Results

Charge of Workshop

Kim McAllister, Ph. D.
**National Institute of Environmental Health
Sciences**

February 6-7, 2012



“Input on Strategies to Encourage Broad Data Sharing in Environmental Health Sciences Research”

Purpose of RFI:

To gather information/recommendations from all stakeholders regarding approaches/strategies that allow broad data sharing in the field of environmental health sciences in human population studies.

Specific Questions of RFI:

- What unique considerations exist for data sharing for studies with environmental exposure data?
- What challenges or barriers exist for researchers wishing to more broadly share their data with others?
- What additional tools or resources do researchers believe will allow more efficient and effective data sharing in the environmental health science community?

Broad Themes Emphasized in RFI Responses:

- Protection of Privacy/Confidentiality Issues
- Institutional Review Board (IRB) issues
- Legal and Regulatory Issues
- NIH Programmatic and Logistical Considerations
- Computational Challenges

Protection of Privacy and Confidentiality:

- ❖ Importance of community-based participatory research, especially in vulnerable communities disproportionately affected by exposures
- ❖ Concerns regarding specific location identification of exposures due to potential discrimination (ex. report of lead paint exposures by locations to departments of health)
- ❖ “Anonomize” data:
 - unique subject identifier
 - Archive data in separate external repository
- ❖ Security is inadequate for online databases

Institutional Review Boards (IRB):

The lack of continuity, consistency, and clarity across IRBs was emphasized by many research groups as a disincentive for attempting to more broadly share their data with others.

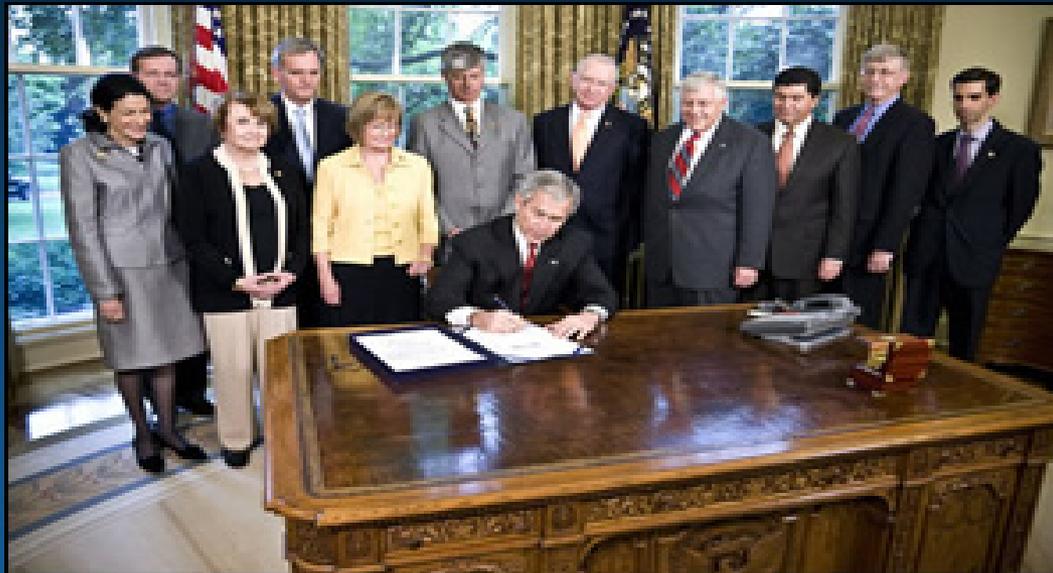
- ❑ Informed consent models need to be redeveloped to allow more sharing
- ❑ Some medical data, protected under HIPPA (Health Insurance Portability and Accountability Act) requirements, may be restricted from sharing
- ❑ Researchers of clinical trials must follow the data safety monitoring boards' recommendations under the guidelines from the FDA as well

Legal and Regulatory Concerns:

Exposure data will continue to be of high interest to regulatory agencies with respect to the evaluation of the health implications of chemicals:

- ❑ The reanalysis and/or reinterpretation of environmental health science data in an effort to delay regulatory reform or influence court cases and the general public.

Genomics has GINA:



President George W. Bush signs H.R. 493, the Genetic Information Nondiscrimination Act of 2008, Wednesday, May 21, 2008, in the Oval Office. White House photo by Eric Draper.

Does exposure data need comparable protection?

Logistical Recommendations for NIEHS:

- ❑ Creation and long-term support of searchable data and sample repositories.
- ❑ Requirement for large collaborative projects to release data into a centralized web-based database.
- ❑ Guidelines for uniformly collected pooled datasets to be securely accessible to many users (ex. *National Database for Autism Research, NDAR*).

Computational Challenges:

- “*Analysis, not data creation, will be the fundamental hurdle preventing further advances in the field of Environmental Health*”.
- Massively parallel data analysis tools with data sharing networks and cloud (or grid) computing cyber-infrastructure will be emerging ideal systems to work towards.
- Data harmonization efforts need standardized measures of exposure (ex. *PhenX, eMERGE, GENEVA, etc.*)

Other Considerations:

- ❑ International collaborations involving multiple foreign institutions may need special considerations
- ❑ Electronic health registry information linked to disease outcomes is a greatly underutilized resource (ex. *Kaiser Permanente*)
- ❑ Potential for subjective re-analysis of epidemiological datasets to “prove” specific hypotheses or inappropriate use of datasets by investigators unfamiliar with the details of the population study

For GWAS Data:

Due to effort, cost, and consortium nature of genome-wide association studies, NIH decided GWAS data was a

- ❖ “community resource project”
- ❖ “founded on the principle of no-cost, rapid, and complete release ...for use by investigators throughout the global scientific community”

Should environmental exposure data also be considered community resource due to valuable nature of data?



National Children's Study: Model for Data Sharing

IMPROVING THE HEALTH
OF AMERICA'S CHILDREN



Key Workshop Questions:

- ❑ What can NIEHS do to facilitate data sharing efforts?
- ❑ Is there a need to identify “best practices” for sharing environmental health data?
- ❑ If NIEHS adopts data sharing guidelines, what are the minimal data requirements that should be included?