Approaches to Streamlining the Data Access Process:

"Policy/ELSI"

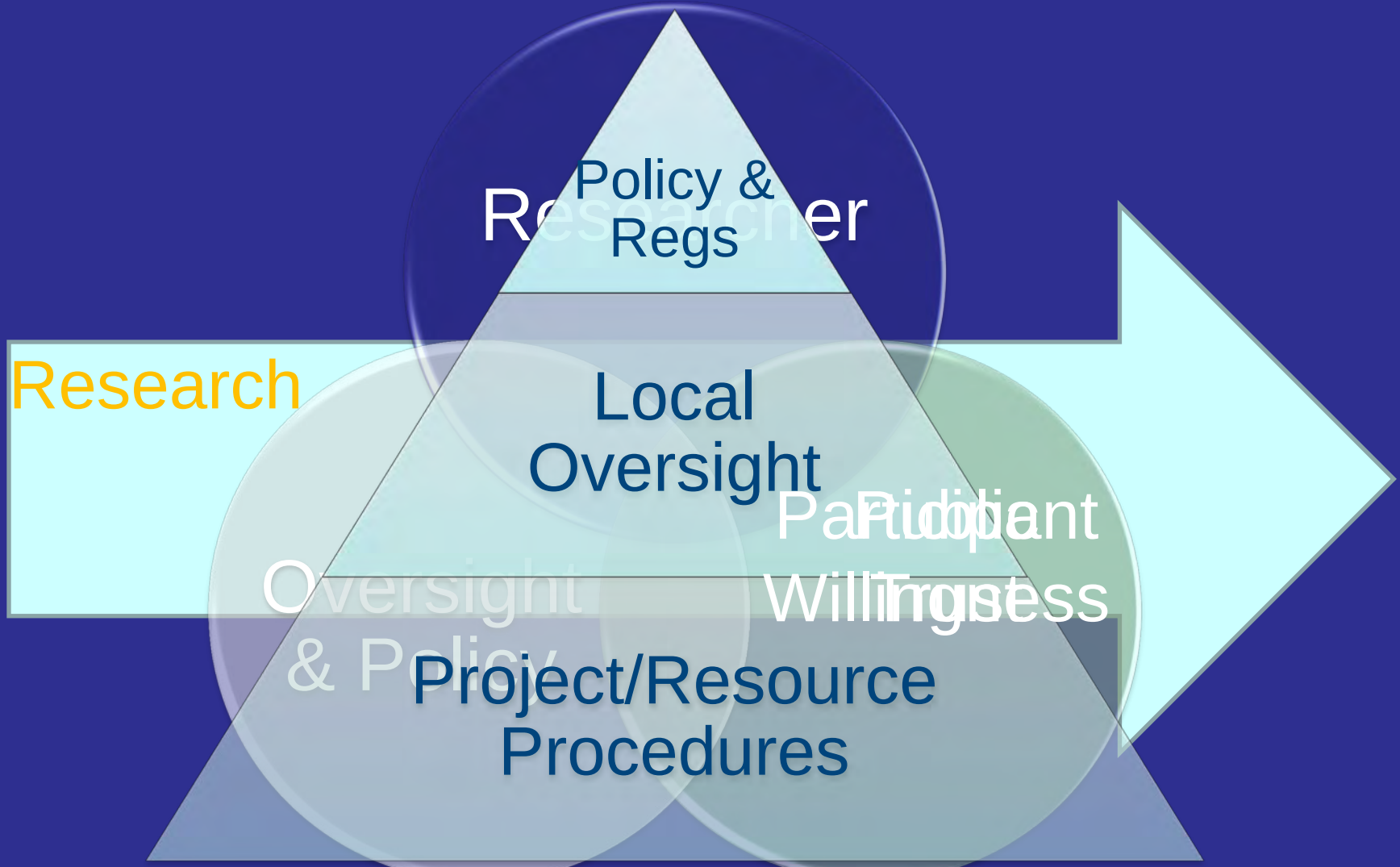
Laura Lyman Rodriguez, Ph.D.
Office of Policy, Communications and Education
National Human Genome Research Institute

Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects
June 5, 2012

The Scientific Aims (Simplified)

- Build data resource(s) with sufficient power and flexibility to ask big questions and find small answers
- Enable necessary high-powered statistical analyses and biologically relevant science
- Work across disease disciplines analyzing multiple data types using common (or at least comparable) “structures”

Systems of Trust/Oversight - *a proxy for ethics*



Considerations

- What are the (priority) scientific aims?
- What ethics or policy issues are raised by each of the models?
  - Participant autonomy – informed consent
  - Participant privacy interests
  - Potential for recontact – additional analysis, return of results
  - Intellectual property issues
- What is possible within existing frameworks versus what would require new policy development?
- How would proposed options scale and track at the national or international level (technical & policy)?

A Shifting Spectrum of "Identifiability"

OPEN ACCESS Freely available online

PLoS GENETICS

Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SN

Nils

John

1 Transl

States

TECHNICAL REPORTS

nature genetics

Ba

fro

Eric

The American Journal of Human Genetics 90, 591–598, April 6, 2012

ARTICLE

On Sharing Quantitative Trait GWAS Results in an Era of Multiple-omics Data and the Limits of Genomic Privacy

Hae Kyung Im,[1,*] Eric R. Gamazon,[2] Dan L. Nicolae,[2,3,4] and Nancy J. Cox[2,3,*]

Recent advances in genome-scale, system-level measurements of quantitative phenotypes (transcriptome, metabolome, and proteome) promise to yield unprecedented biological insights. In this environment, broad dissemination of results from genome-wide association studies (GWASs) or deep-sequencing efforts is highly desirable. However, summary results from case-control studies (allele frequencies) have been withdrawn from public access because it has been shown that they can be used for inferring participation in a study if the individual's genotype is available. A natural question that follows is how much private information is contained in summary results from quantitative trait GWAS such as regression coefficients or p values. We show that regression coefficients for many SNPs can reveal the person's participation and for participants his or her phenotype with high accuracy. Our power calculations show that regression coefficients contain as much information on individuals as allele frequencies do, if the person's phenotype is rather extreme or if multiple phenotypes are available as has been increasingly facilitated by the use of multiple-omics data sets. These findings emphasize the need to devise a mechanism that allows data sharing that will facilitate scientific progress without sacrificing privacy protection.

Updating the Common Rule

- Genetic samples considered inherently identifiable – risk classified as “informational”
- Data security protections, calibrated to level of identifiability
- Written consent required for all uses of existing research samples (short forms, broad consent OK; only applies prospectively)

Advanced Notice of Proposed Rule Making (ANPRM)

Read more about the July 22, 2011 ANPRM for changes under consideration to the Common Rule.

These changes, the most extensive since the Department of Health, Education, and Welfare published proposed rules for the protection of human subjects involved in research on August 14, 1979, are **available for public comment until September 26, 2011**.

Presidential Bioethics Commission

Among other issues, the Commission is interested in receiving comments on:

- The implications of large-scale human genome sequencing for the privacy of individuals, research subjects, patients and their families;
- The views of those groups and medical professional communities about privacy, both as regards genomic information and evolving notions of privacy, as evidenced and influenced by social media; and
- Models and mechanisms for protecting privacy, in both genetic/genomic databases and biobanks, but also in large databases of sensitive information.

The Commission is further interested in receiving comments on:

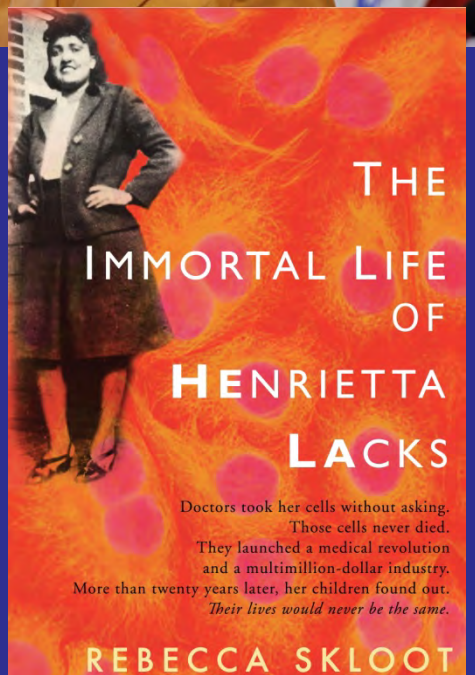
- Issues related to balancing individual and societal interests with regard to the sharing of and access to large-scale human genomic data;
- The views of patients and other stakeholders on who should have access to these data and who should control access;
- Models and mechanisms for governing access to genomic information;
- The role of health information technology in providing and governing access to genomic data; and
- Access to genetic/genomic information by law enforcement entities.

The View from the Public

Indian Tribe Wins Fight to Limit Resea

Edmond Tilousi, 56, who can climb the

By AMY HARMON
Published: April 21, 2010



Data Access Models for Today

- Open data access
- Streamlined controlled access
- Certified researchers and a research commons
- Central analysis server groups

Autonomy and Consent

- What are participants willing to “let” their data be used for?
- What are the limits of “informed consent”
- What are realities of the research paradigm?

Autonomy and Consent

Open Access: Can participants (or scientists) grasp the implications? Will costs include loss of some populations/individuals in the data resource?

Streamlined Controlled Access: Where can efficiencies to process built around consent be created? Broad and open consent? What governance is needed?

Research Commons: Who certifies? Oversight and monitoring of certification and of use? Reciprocity

Central Analysis: How can data use limitations attach to data for the type of broad based analysis envisioned?

Privacy: What are the risks?

- Assumed and accepted in Open Access
- Can they be articulated? quantified?
- Streamlined, Research Commons, and Central Server models attempt to manage – is that possible?
- Long-term: assess the trade-offs for individuals and for society

Recontact and Intellectual Property

- Recontact is ideal, but will necessitate new policy work
  - Research and experience needed
- IP principles vary across research sectors, establish baselines

Governance

- Intrinsic to various models are promises of conduct
  - to maintain trust must include reasonable expectation for compliance
- Issues of scale and practicability

Streamlined Controlled Access Model

- Improvements being considered:
  - Building a standard lexicon of Data Use Limitations to increase transparency and consistency
  - Piloting centralized review for multi-study requests
  - Simplified access process for aggregate data for some data sets
  - Filters within dbGaP to enhance the ability of users to find relevant data

Finding the balance …

… assessing, learning, adjusting

Laura.Rodriguez@nih.gov

genome.gov