



KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

CENTER FOR DEMOCRACY
& TECHNOLOGY

Distributed approaches to health data analysis

Harley Geiger
Policy Counsel, CDT
March 2, 2012



Secondary use programs

- ...are a good thing!
 - But should maximize privacy and security.
- Current trend is to use centralized databases to collect health claims for analysis.
 - Regulations are locking plans into this model.
 - Examples:
 - CMS' proposed risk adjustment program
 - State all-payer claims databases (APCDs)



Problems with centralization

- Data breaches
- Public trust
- Inefficient, costly, and burdensome
- Scope creep
- Poor long-term strategy



The decentralized alternative

- Provide researchers and agencies with the results of their analyses, but not copies of the raw data.
 - Reduces risk of data breach
 - More in line with public privacy expectations
 - Eases plans' proprietary concerns
 - Minimizes data transfer
 - Leverages existing infrastructure and expertise



Distributed query systems

- Researchers write analytic code and send the code to payers
- Payers analyze their in-house data and provide researchers with the results
- No common data format required
- Example: FDA's Mini-Sentinel
- Can be used for many research purposes
- May not be appropriate for all health reform goals



Distributed *access* systems

- Payers set aside data in secure environment and provide researchers with secure access
- Researchers analyze data in the secure environment and keep the results, not data copies
- Common data format is required
- Can be used for secondary use programs that would provide a competitive edge to plans
- Technical challenges must be overcome



Recommendations

- Keep regulatory language flexible enough to preserve possibility of non-centralized models in the future (subject to approval)
- Explore distributed solutions for secondary use programs



Thank you

harley @ cdt . org

@HarleyGeiger