Learning from Big Clinical Data

Resources and Roles for Public Management Scholars in the Big Clinical Data Realm

Introduction

- Sara Jordan: Center for Public Administration and Policy, Virginia Tech
  - srjordan@vt.edu
- Big Data concern: Ethics, Global regulation
- Clinical Data concern: Clinical trials registries, Research misconduct
- Big Clinical Data concern:
  - How do clinical trials registries ensure compliance with international regulations on the protection of human participants and their identifiable data when used in Big Data initiatives?

Two Purposes

Big Clinical Data Resources

- What Big Data options are available for clinical applications?
  - Governmental, specialist and general
  - Industry, specialized and general

Public Management Research Role

- Which Big Data options should we invest in as public policy researchers?
  - A few steps for judging “good” Big Clinical Data

Seven V's of Big Data

1. Volume: amount of data
2. Velocity: speed according to which data can be gathered or analyzed
3. Variety: subject types and sources of data that can be aggregated
4. Veracity: verifiable pedigree of data
5. Value: relationship to tractable problems or profit-making solutions
6. Valid: integrity of the data guaranteed for fitness for purpose and consistency in application over the data lifecycle
7. Voluntary: data is volunteered by users and algorithms are written by public good motivated "white-hat" hackers*

Big *Clinical* Data

Big Health Data

- High volume, velocity, and variety data, analyzable via structured or unstructured inquiry to develop hypotheses about the determinants of overall health and well being
  - Personal fitness data from the “internet of things”
  - Health administration data such as the NHANES or NHIS

Big Clinical Data

- High value, veracity, variety data validated to generate hypotheses and make predictions about the clinical efficacy of tested interventions
  - Clinical trials registries
  - OpenfMRI database
  - - “Omics” data

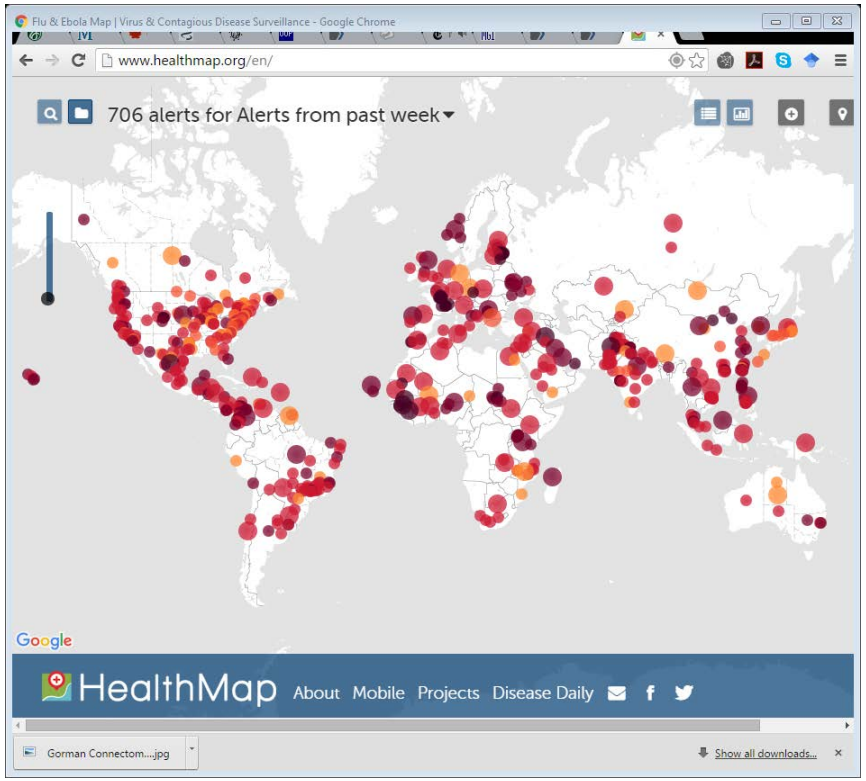
Big Clinical Data

- Image/ Health: high-near term-value, high-validity, fast data requiring maximum security due to strong relationship to identifiable persons
  - Example: Electronic Health Records (EMR) proposed by Flatiron, inc.
  - Example: GoogleFlu

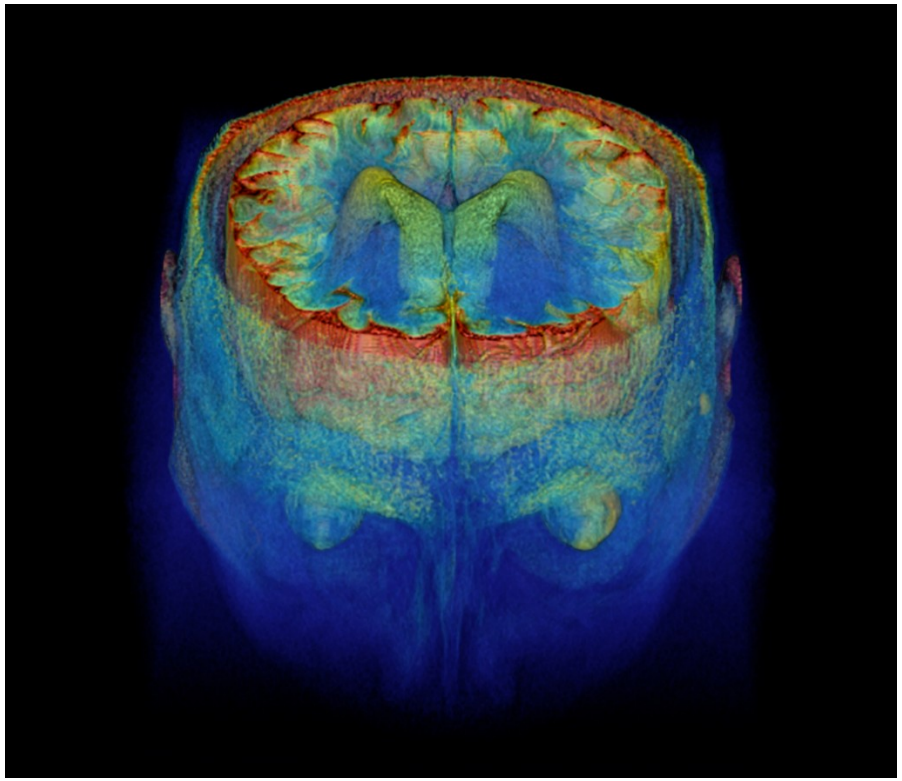
- Reality/ Clinical: unknown-near or long term value- moderate validity, slow data that bears probabilistic resemblance to identifiable individuals
  - Example: -Omics databases
  - Example: Clinicaltrials.gov

Big Health and Big Clinical Data

Healthmap Project



Human Connectome Project



Finding Big Clinical Data

Governmental

- Know what you need before wading in!
  - NIH Data-Sharing Repository, National Library of Medicine
  - PubMed, National Library of Medicine
  - Clinicaltrials.gov
  - WHO: International Clinical Trials Registry Platform

Non-Governmental

- Be prepared to spend time and money!
  - Non-governmental organizations
    - Cochrane Reviews
  - Industry
    - Flatiron Health
    - PhRMA member adherence to data sharing principles

Big Clinical Data Investment Strategy

- Given the:
  - Substantial barriers to entry for use
  - Need for validated data
  - High securitization demands
  - Potential commercialization of results
- What strategy could a public policy scholar use to determine data in which to invest?
  - Accessibility
  - Durability

Big Clinical Data Access

- Restricted use and registered user databases are common in for clinical applications
  - Concerns about de-anonymization of individuals through merging databases
  - Barrier to entry for professional use only
  - Embargoes due to publication and grant requirements
- Government data will require registration in most cases
- Industry data will require a research proposal in some cases
  - There are “back door” entries from github to clinical data, but these have validation issues

Big Clinical Data Evaluation for Durability

Collaborative Governance

- Initiation and/or sponsorship by public agency or agencies
- Multi-sector participation
- Direct collaborator participation
- Formal organization and collective action
- Consensus based decision-making
- Public policy focus
  - Ansell and Gash (2008, 544-545)

Data Governance

- Clear directions for structuring submitted data or for structuring webpage information
- Clear requirements for data sharing and use
- Repository for analysis and findings

Durable Big Clinical Data

- European Clinical Research Infrastructures Network standards for GCP-compliant data:
  - Information technology (IT) infrastructure & culture
  - IT support
  - Clinical Data Management applications
  - Trial Data management
  - International Compliance

Durable Big Clinical Data (-)

- Expect IP issues to intervene on data openness
  - Download or scrape data in to a structured format
- Expect de-anonymization attacks to restrict data access
  - Higher registration and proposal barriers to entry
  - More “limited” data
- Expect changing privacy laws and “right to be forgotten” policies to restrict access
  - Monitor use stipulations– it’s your responsibility, not theirs

Durable Big Clinical Data (+)

- More multi-site and virtual clinical trials collaborations will mean more data shared on the web
- Changes in informed consent processes, language, and requirements may mean anonymization is of less concern in future trials
- Data sharing norms are changing for the better
- Inclusion of GIS coding in clinical trials is increasing

Resources (Links to Data)

- Clinicaltrials.gov: main resource for clinical trials information and results data for studies that must comply with FDA rules
- NIDA Clinical Trials Network database: Data on drug abuse trials and related studies: https://datashare.nida.nih.gov/
- ACCT or Aggregate Analysis of ClinicalTrials.gov data set from the Clinical Trials Transformation Initiative: http://ctti-clinicaltrials.org/
- World Health Organization: International Clinical Trials Registry Platform: http://apps.who.int/trialsearch/AdvSearch.aspx

Resources (Links to Ethics and Laws)

- ICJME: “Clinical Trials Registration”
- PhRMA: “Principles for Responsible Clinical Trial Data Sharing”
- FDAAA 801: FDAAA Section 801