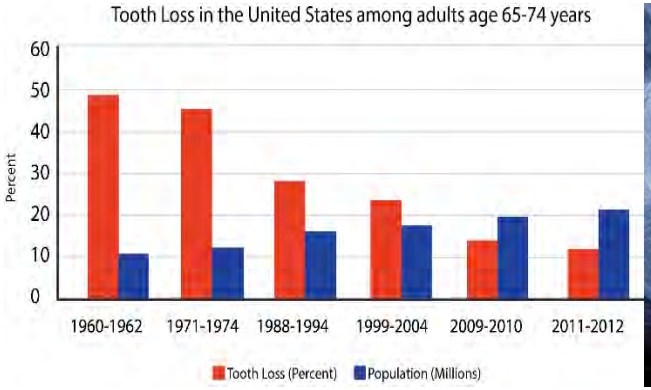
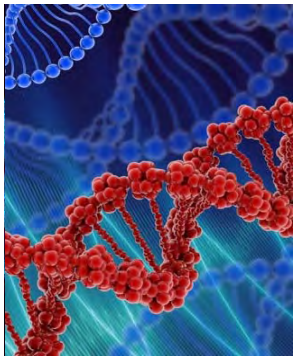
National Institute of Dental and Craniofacial Research

BIG DATA and Oral Health Research

Bruce A. Dye, DDS, MPH

19 April 2016





DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

Informatics-BIG DATA-Oral Health Research

The Landscape Ahead

- Brief Introduction to Health Informatics
- Brief Description of what BIG DATA is
- Highlight some BIG DATA projects using publicly available data

Informatics-BIG DATA-Oral Health Research

NIH National Institute of Dental and Craniofacial Research

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Informatics Science is one of the fastest growing research fields
- It is a multidisciplinary approach to information and knowledge management in biomedical research, clinical care, and public health
- Origins in computer science / information technology
- Evolving with more of a focus on Data Science

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Informatics is Transforming the Way We Work
  - Data mining,
  - Natural language / text processing,
  - Cognitive science,
  - Clinical research,
  - Genomics / proteomics, and
  - Database management and data analytics for large amounts of data generated in health care delivery or produced through public health activities.

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Recognized Core Areas of Informatics (AMIA)
  - Translational Bioinformatics
  - Clinical Research Informatics
  - Clinical Informatics
  - Consumer Health Informatics
  - Public Health Informatics

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Why is BIG DATA becoming a big deal in Informatics?
  - Facilitating biomedical science
  - Promoting the quality and safety of patient care
  - Improving population health

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- BIG DATA can be characterized by:
  - The amount of collected data (Volume)
  - The source and type of data (Variety)
  - The speed of generated data (Velocity)
  - The quality of incoming data (Veracity)

Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

- Data Accumulation Matters (Velocity)
  - Data accumulates expansively
  - No economic trade off between retaining vs deleting data
- Data Quality Matters (Veracity)
  - More data doesn't always give us more "value"
  - Information from data becomes more valuable when the data is more reliable
  - Data-driven decision-making requires accurate and reliable data

Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

- Data Size Matters (Volume)
  - At its most elementary level—BIG DATA is about bringing datasets together
  - There is no critical mass of data alone needed to make it "Big."
- Data Complexity Matters (Variety)
  - You need a mix of varying types and sources of data to make it complex and large.
  - Data linkages are important

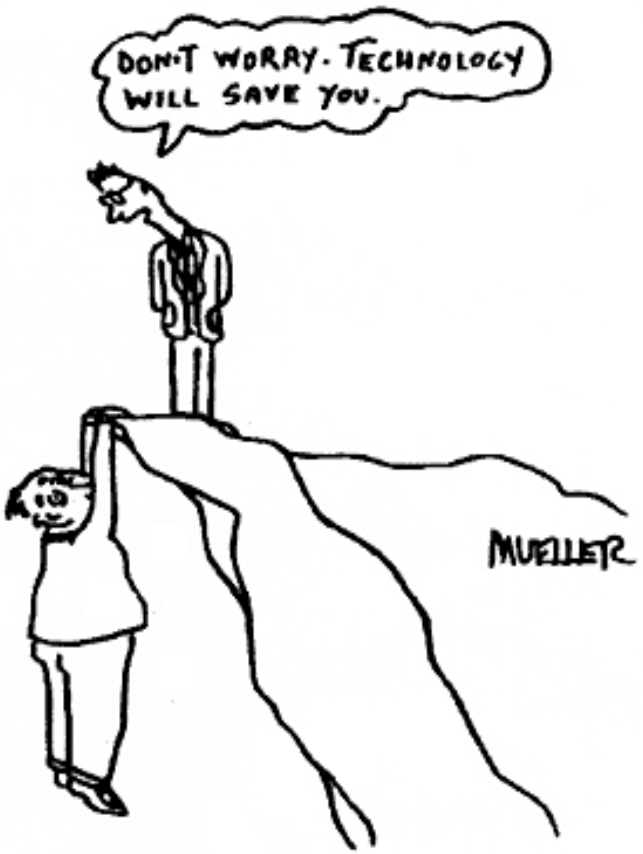
Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

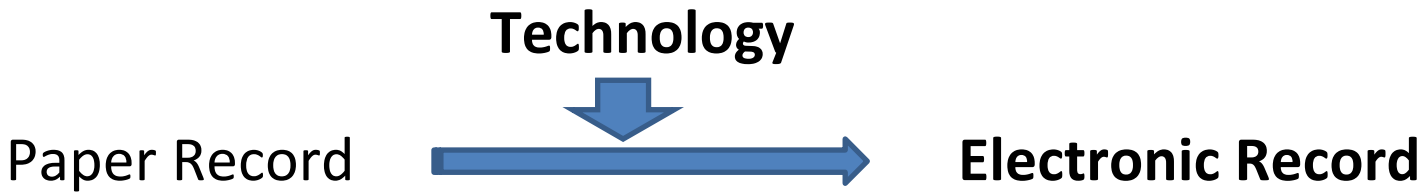
•"Big Data" begins to form when a group of data sets are brought together become so large and complex that it begins to challenge contemporary data processing and analytical approaches.

Informatics-BIG DATA-Oral Health Research

- Technology is also propelling this phenomena of “Big Data” and it’s use in research.
- Technology is allowing us to collect, process, and share data easier and faster then ever before.



Informatics-BIG DATA-Oral Health Research



- Technology and creation of Electronic Health Record (EHR).
- Big Data analytics are not about analyzing a provider's EHRs in isolation.
  - For example, to identify how many widgets were rendered
- It is about analyzing provider health recorders to identify factors that are **associated** with those widgets rendered.

Informatics-BIG DATA-Oral Health Research

- The contribution of technology to the evolution of BIG DATA
  - described well by Kenneth Cukier (TED talk).
  - "We use to look at small data to try to understand the world."
  - "The idea is that more data doesn't let us see more of the same but it allows us to see better, to see different, to see something new."
- Inherent to this is that BIG DATA methods permits us to observe varying amounts of information in a way that we can extract value faster and more meaningful.

Informatics-BIG DATA-Oral Health Research

- Observation is fundamental to Epidemiology.
  - Epidemiology is the study of the distribution and determinants of diseases and conditions in man.
  - Epidemiology is essentially all about collecting data and analyzing it to reveal patterns, trends, and associations.
- BIG DATA is about observing better—identifying patterns in ways that make us better understand the world we live in.

What is “Big Data”?

BIG DATA is:

Very large complex datasets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human health, behavior, and interactions.

NIH National Institute of Dental and Craniofacial Research

Examples of BIG DATA Projects

- Using linked data for BIG DATA research projects

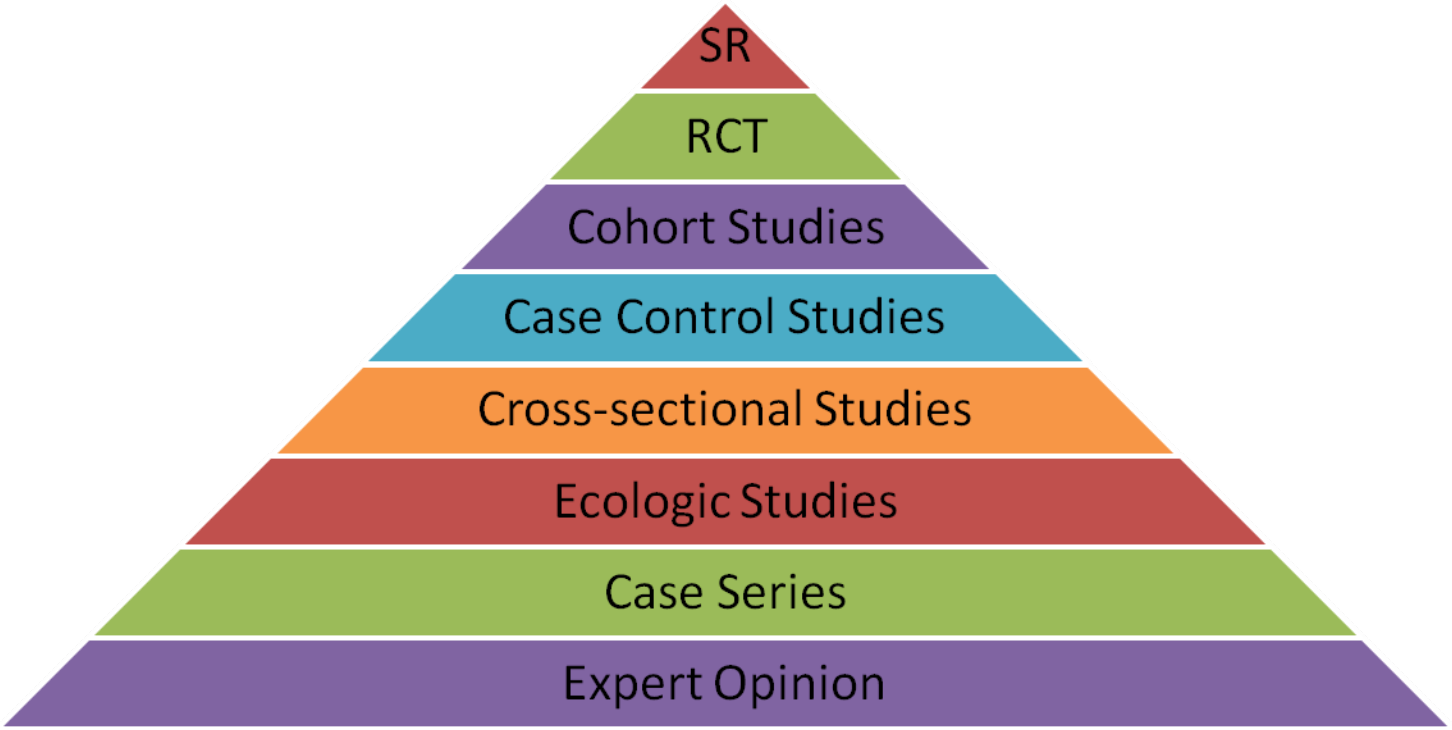
NIH National Institute of Dental and Craniofacial Research

Examples of BIG DATA Projects

NIH National Institute of Dental and Craniofacial Research

Examples of BIG DATA Projects

- Review Study Design Heirarchy & Levels of Evidence

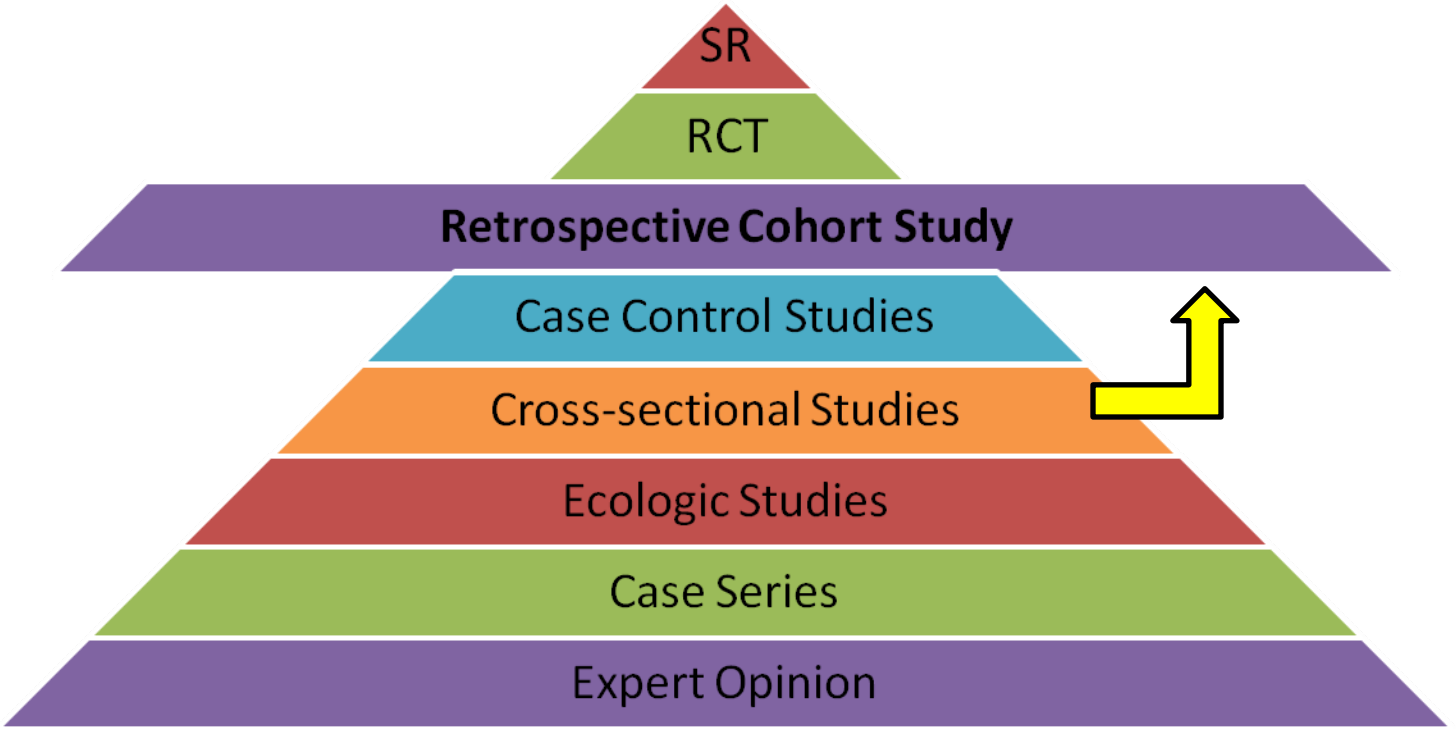


Examples of BIG DATA Projects

- What is the potential of BIG DATA?
- Observational studies on steroids
  - Where the sum has **more added value** than the individual data files

Examples of BIG DATA Projects

- Boosting NHANES from a good x-sectional study to a good cohort study accounting for temporal influence

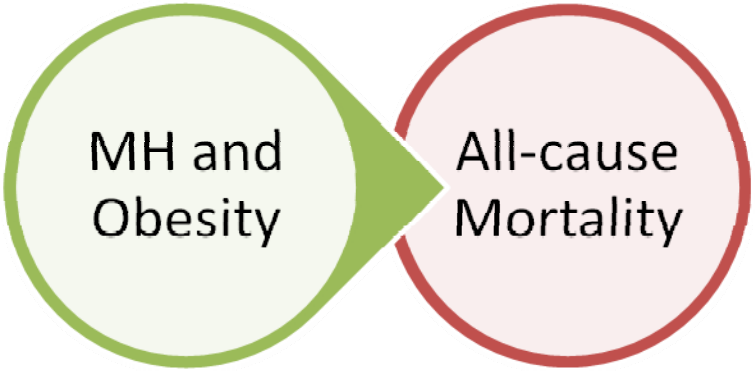


Examples of BIG DATA Projects

- Example 1: NHANES and Death Records

Examples of BIG DATA Projects

- Obesity really associated with increase risk of death?
- NHANES contributes baseline information
  - Age 17-60 in 1988-1994
  - Body Measures: Obese ≥ 30 BMI
  - Lab Profiles: Metabolic Health
- NDI contributes mortality information:
  - Underlying and multiple causes of death through 2006



Examples of BIG DATA Projects

- Compared to healthy lean group
  - Risk of all-cause mortality was twice as high in Metabolic unhealthy obese group
  - No risk difference in metabolic healthy obese group
- Strengths
  - Nationally representative sample of the US population
  - Robust end point (all-cause mortality)
  - Relatively long follow-up period (~15 years)
  - Measured rather than self-reported weight and height
  - Data collection is standardized and use appropriate quality controls.

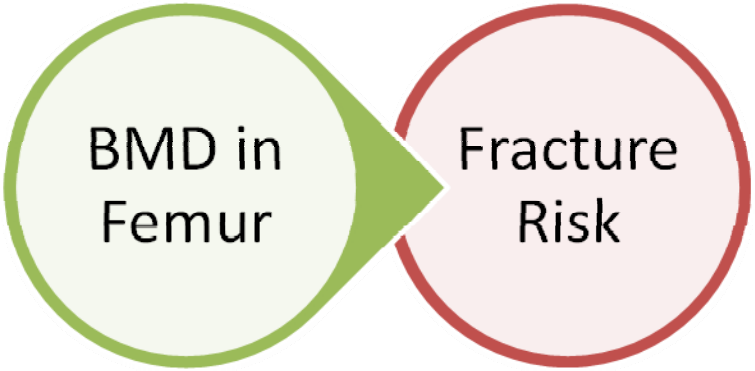
Durward CM1, Hartman TJ, Nickols-Richardson SM. All-cause mortality risk of metabolically healthy obese individuals in NHANES III. J Obes. 2012; doi: 10.1155/2012/460321.

Examples of BIG DATA Projects

- Example 2: NHANES and Medicare Records

Examples of BIG DATA Projects

- Bone mineral density really associated with increase risk of fracture?
- NHANES contributes baseline information
  - Bone Mineral Density in Femur (DEXA) in 1988-1994
  - Age 65 and older
- Medicare contributes event information:
  - Incident fracture from Medicare Claims Files through 2007



NIH National Institute of Dental and Craniofacial Research

Examples of BIG DATA Projects

- Compared to healthy lean group
  - For each SD decrease in BMD: risk of fractured doubled
  - Age, sex, R/E was still associated with increase risk of fracture even after controlling for femur BMD
- Strengths
  - Nationally representative sample of the US population
  - Clinically diagnosed end point (ICD-9)
  - At least a decade-long follow-up period
  - Data collection was standardized and used appropriate quality controls.

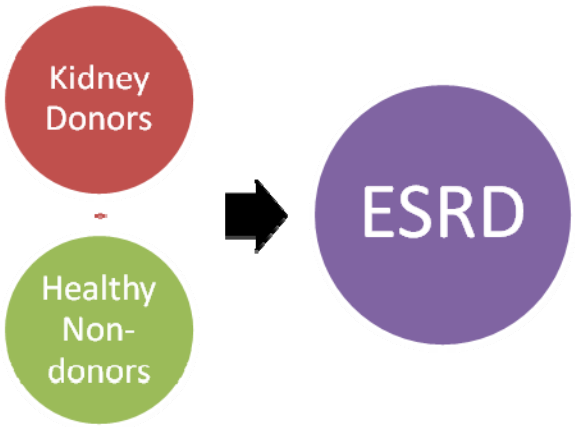
Looker AC. Femur neck bone mineral density and fracture risk by age, sex, and race or Hispanic origin in older US adults from NHANES III. Arch Osteoporos. 2013. doi: 10.1007/s11657-013-0141-4.

Examples of BIG DATA Projects

- Example 3: NHANES, USRDS and CMS Data

Examples of BIG DATA Projects

- Greater risk of ESRD after live kidney donation?
- US Renal Data System
  - Registry of all kidney donations (~96,000 from 1994-2011)
  - Identify the SES, BMI, Smoking and Blood Pressure in Cases
- CMS information identifies ESRD
  - CMS Medical Evidence Form 2728—Certification of ESRD
- NHANES information creates healthy non-donor group:
  - Data 1988-1994 used to create population control
  - Replacement Matching use SES, BMI, Smoking and Blood Pressure information to build a control group of ~96,000 people



NIH National Institute of Dental and Craniofacial Research

Examples of BIG DATA Projects

- Estimated Risk of ESRD at 15 Years of DOnation
  - 30.8 / 10,000 for donors
  - 3.9 / 10,000 for healthy non-donors
- Strengths
  - Inclusion of all kidney donors in 20 year period
  - Use of data from people not selected based on existing health condition to create healthy non-donor pool
  - Availability of clinically diagnosed end points
  - A ~15 year follow-up period
  - Data collection was standardized and used appropriate quality controls.

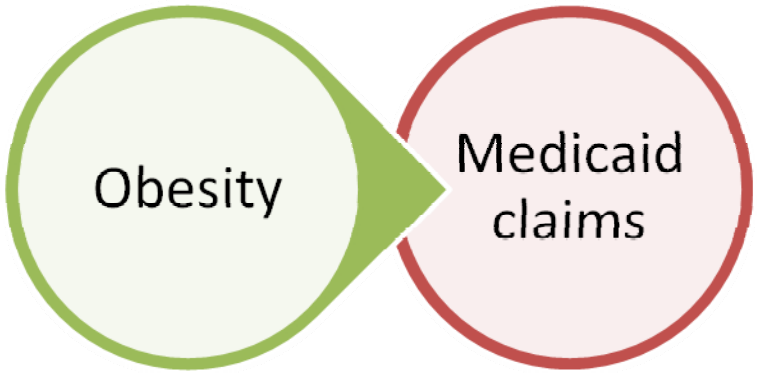
Muzaale AD, Massie AB, Wang MC, Montgomery RA, McBride MA, Wainright JL, Segev DL. Risk of end-stage renal disease following live kidney donation. JAMA. 2014 Feb 12;311(6):579-86.

Examples of BIG DATA Projects

- Example 4: NHANES and Medicaid Claims Data

Examples of BIG DATA Projects

- What is the econometric cost of obesity to Medicaid?
- NHANES contributes baseline information
  - Age 20+ in 1999-2004
  - Body Measures: Obese ≥ 30 BMI
- Medicaid Analytic eXtract (MAX) data contributes:
  - Amount of FFS costs paid through Medicaid same year



Examples of BIG DATA Projects

- Average FFS costs paid by Medicaid in same year as obesity was determined during 1999-2004
  - $95.02 for non-obese Medicaid receiptents
  - $225.72 for obese Medicaid receiptants
- Strengths
  - Nationally representative sample of the US population
  - Measured rather than self-reported weight and height
  - Actual costs paid based on claims data
  - Data collection was standardized and used appropriate quality controls.

Medicaid Policy Brief by MATHEMATICA January 2013. Allison Dodd and Phillip Gleason authors

Informatics-BIG DATA-Oral Health Research

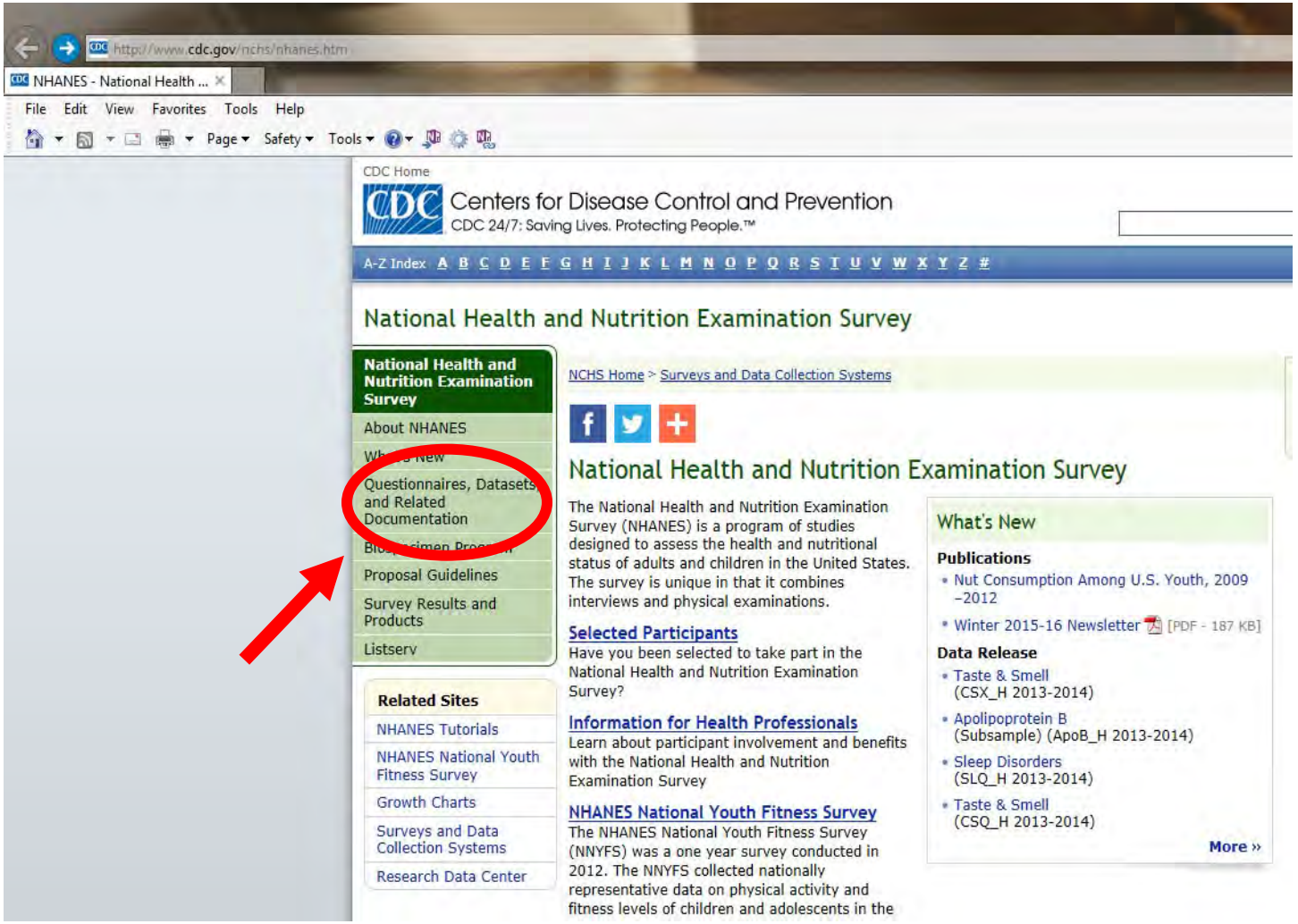
- NHANES and BIG DATA
  - Information from data becomes more valuable when the data is more reliable
    - With NHANES—we have Veracity
  - It's about observing better—identifying patterns in ways that make us better understand the world we live in
    - With Linkage—we obtain Volume, Variety and Velocity
  - Facilitating Data-driven decision-making

Informatics-BIG DATA-Oral Health Research

Thanks!

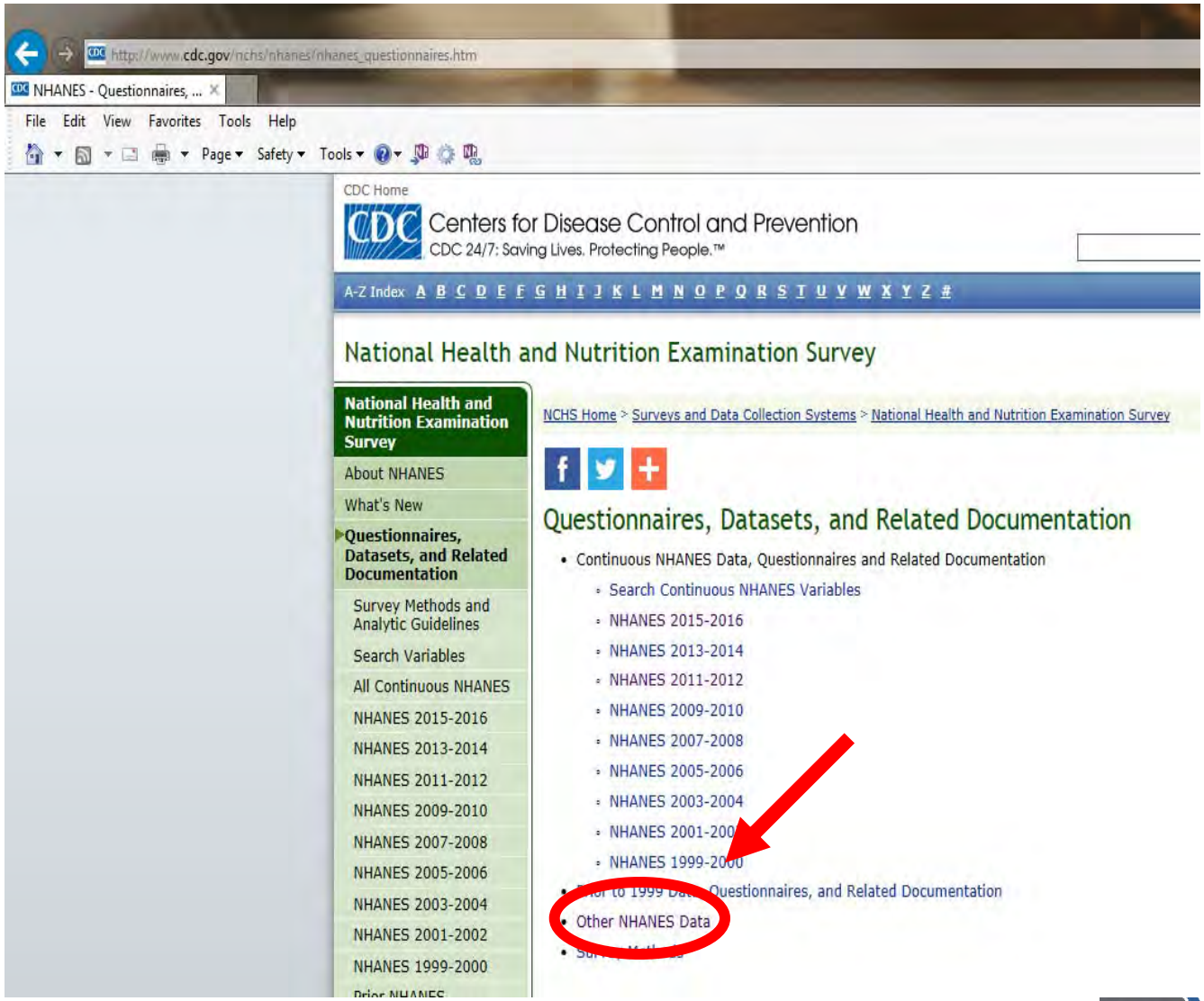
NIH National Institute of Dental and Craniofacial Research

Informatics-BIG DATA-Oral Health Research



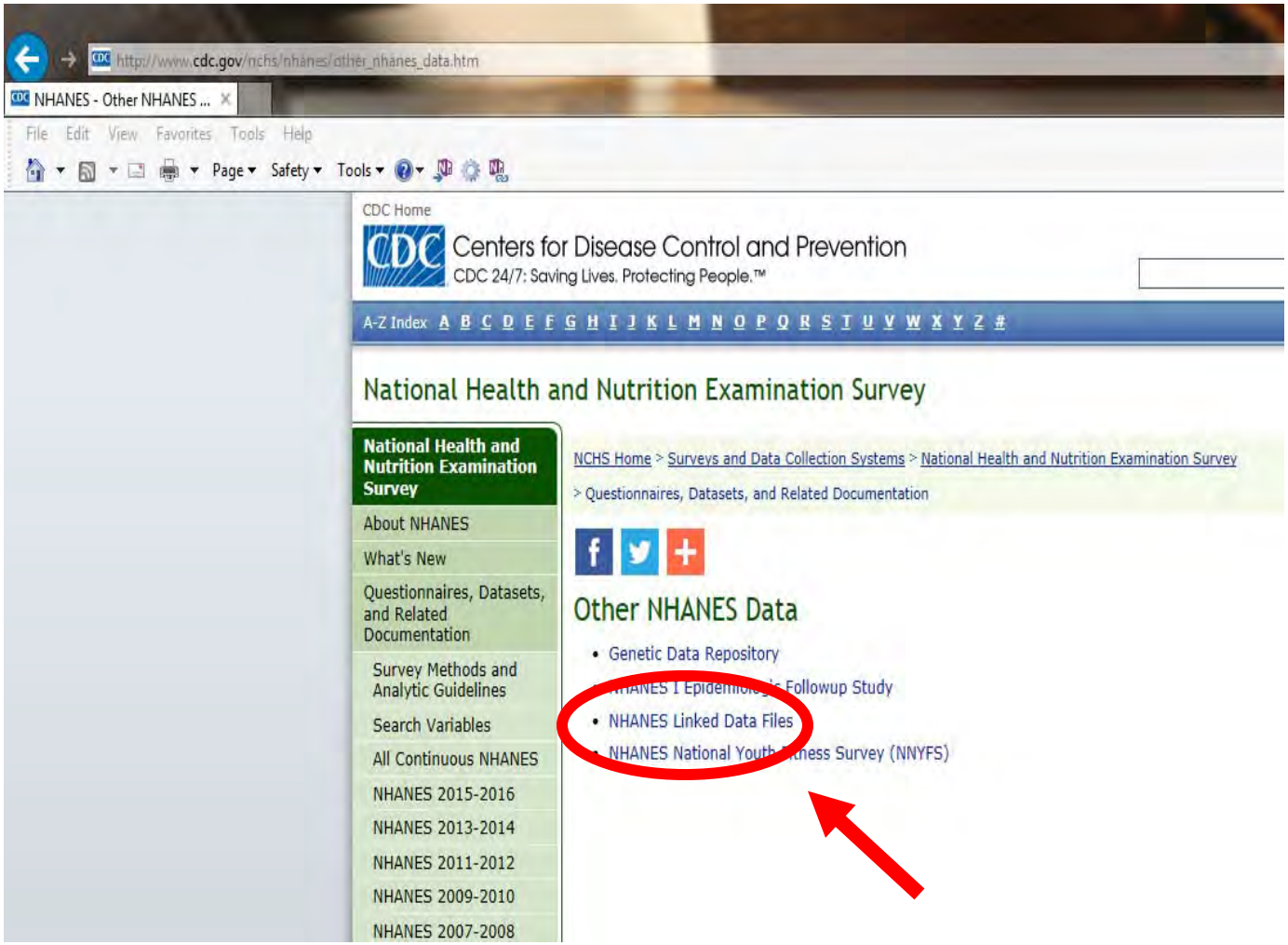
NIH National Institute of Dental and Craniofacial Research

Informatics-BIG DATA-Oral Health Research



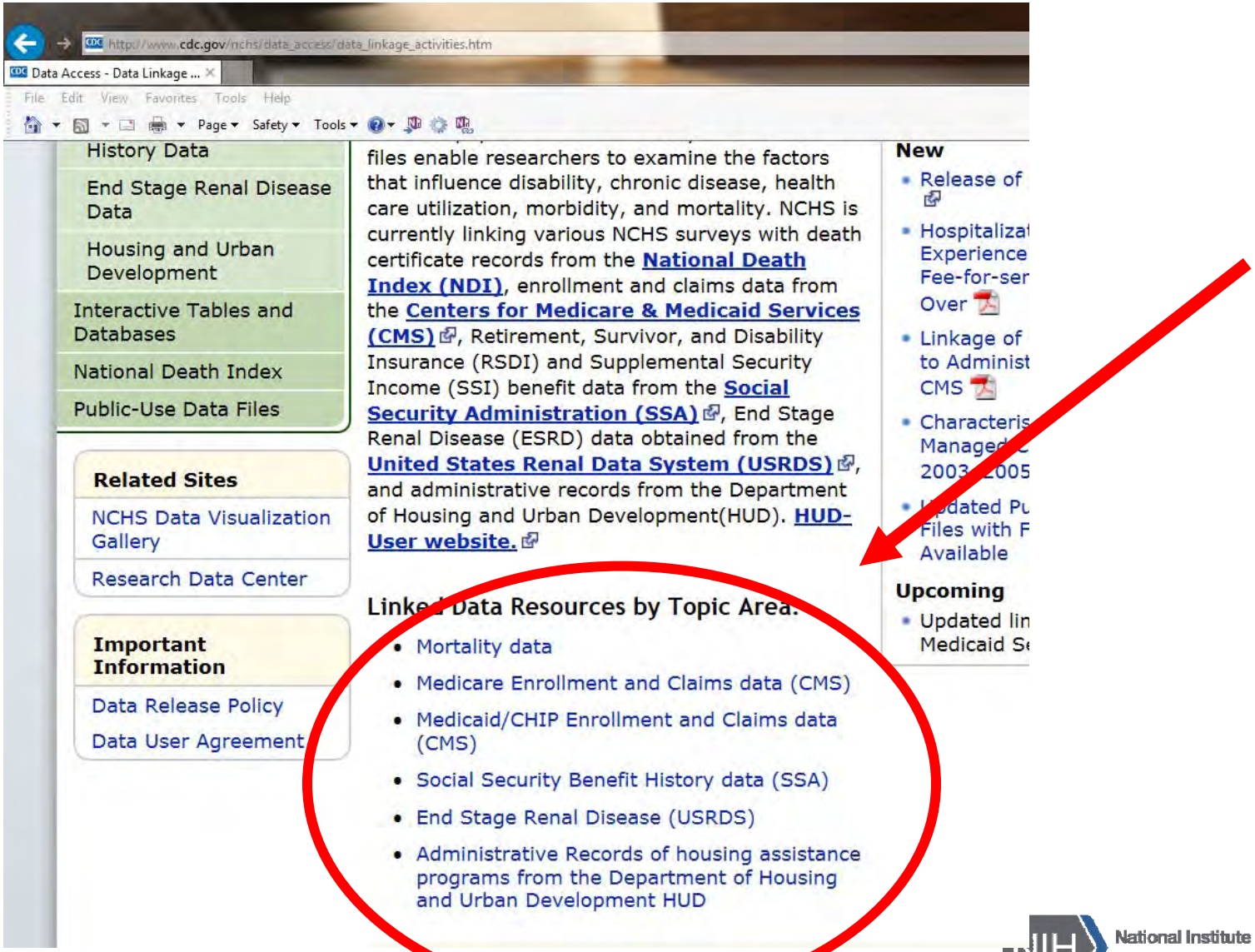
NIH National Institute of Dental and Craniofacial Research

Informatics-BIG DATA-Oral Health Research



NIH National Institute of Dental and Craniofacial Research

Informatics-BIG DATA-Oral Health Research



National Institute of Dental and Craniofacial Research

NIDCR Training Opportunities Big Data Research

Residency in Dental Public Health and Oral Health Informatics (Emerging)

- Partnership with National Library of Medicine (NLM)
- For dentists with an interest in informatics science
- An MPH or equivalent public health-related degree required
- 2-Year Full-time Residency / Fellowship
- Mentors and Faculty come from NLM and NIDCR

NIH National Institute of Dental and Craniofacial Research

NIDCR Training Opportunities Big Data Research

Residency in Dental Public Health and Oral Health Informatics (Emerging)

- Informatics Research Area Strength: Public Health
  - Focus on "Big Data" projects
  - Other Potential Research Areas: Bioinformatics and Clinical Research
- Certificate in DPH
- NIDCR Residency in DPH is CODA accredited
  - Meets education qualifications for graduates to challenge DPH certifying exam
- Certificate in Health Informatics
- Goal is to begin application process Summer 2016

NIH National Institute of Dental and Craniofacial Research

NIDCR Training Opportunities Big Data Research

Questions about Residency in Dental Public Health and Oral Health Informatics:

- Bruce Dye
  - NIDCR Residency Program in Dental Public Health
  - bruce.dye@nih.gov