

# Promoting Research Data Discovery

## Development of a Data Catalog at the University of Maryland, Baltimore

Patricia Hinegardner, MLS

Associate Director for Resources

Na Lin, MLS

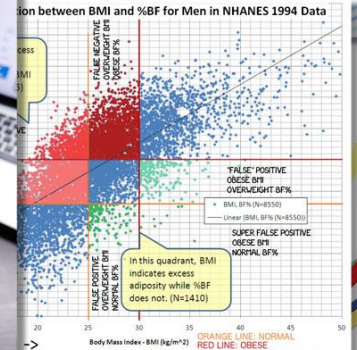
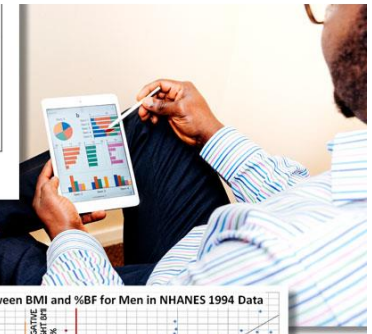
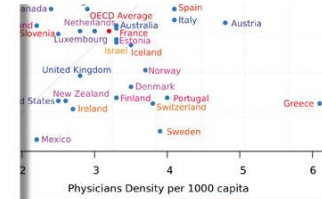
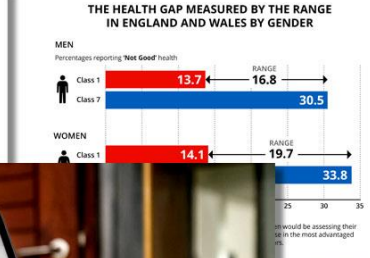
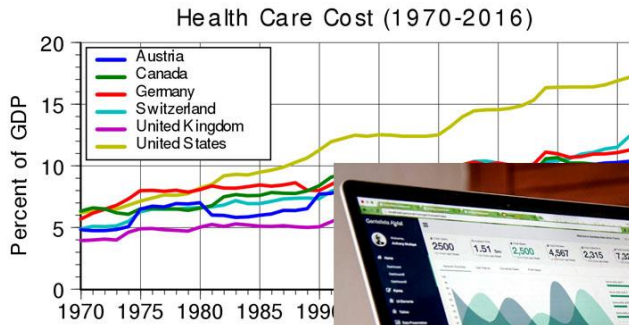
Head, Resource Development and Sharing

**CALD Annual Meeting**

Friday, April 26, 2019

Columbia, Maryland

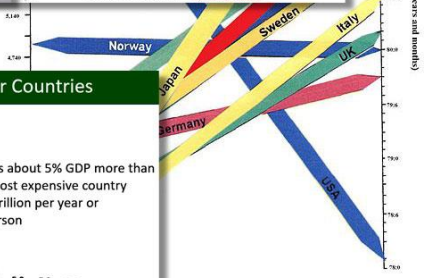
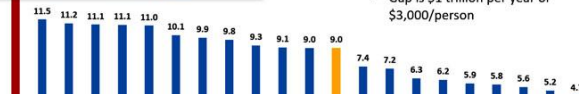
# DATA IS EVERYWHERE!



High Relative to Other Countries

Healthcare Expenses (% GDP)

- U.S. spends about 5% GDP more than the next most expensive country
- Gap is \$1 trillion per year or \$3,000/person



# UMB Data Catalog

A tool to:

- facilitate the discovery of datasets created or used by UMB researchers
- centralize and highlight what is happening at the UMB
- promote interdisciplinary collaboration by identifying common research interest
- support the process of data re-use

# Background

- New York University model - developed software and a metadata schema
- Data Catalog Collaborative Project (DCCP), a multi-university initiative led by New York University
  - Purpose: to exchange ideas
  - Collaborators: UMB, U Pitt, UNC-Chapel Hill, Duke, UVA, Wayne State, Hofstra, Weil Cornell

# HSHSL Data Catalog Team

Project Lead - project management/outreach

Metadata Librarian - creates records

Bioinformationist – outreach

Technology Support - web developer/programmer

Senior Advisor - outreach/promotion

# Technology

- Software available on GitHub - <https://github.com/nyuhsl/data-catalog>
- Operating systems - Windows / Linux
- Software – Symfony2 (Apache & MySQL), Solr
- Web-based interface
- Open Sandbox at <http://dccp-test.hslls.pitt.edu/>

# Outreach Strategies

- Organizational
  - Research deans, program directors
  - Work with IRB and ORD to document UMB data sharing approval process
- Individual Researchers
  - UMB Digital Archive contacts
  - Bioinformationist
  - Research, Education and Outreach librarians for each school

# Outreach Strategies

- Targeted promotion
  - Presentations to heavy data generators
    - Epidemiology
    - Center for Vaccine Development
  - Promotional material tailored to schools
    - handouts
    - news blog posts



# Locating Data

- Databases: NIH Reporter, Pubmed Central, Scopus
- Public repositories: Dryad, Figshare, BioLINCC
- Open access publishers: PLOS ONE, Springer Nature

# Building a Record

- Assign metadata based on article(s) associated with dataset, includes
  - Assign authoritative subject headings (MeSH and LC)
  - Compose description of dataset
- Researcher reviews and approves the draft for publishing

# Protection from Cholera with CVD 103-HgR UMB Dataset

Alternate Title(s): Efficacy of Single-Dose Live Oral Cholera Vaccine CVD 103-HgR

UID: 10

Author(s): Wilbur H. Chen\*, Myron M. Levine  
\* Corresponding Author

## Description

This dataset is the result of a phase 3, randomized, double-blind, placebo-controlled efficacy trial of a single dose live oral cholera vaccine. PXVX0200 was developed to rapidly confer protection to humanitarian relief workers and other travelers visiting cholera-endemic areas as well as residents in regions experiencing outbreaks. Existing licensed vaccines require 2 or 3 (children) doses weeks apart. The trial consisted of 3 groups of volunteers each of which were randomly split into receiving either the PXVX0200 vaccine or a placebo. One group was challenged with virulent *V. cholera* O1 El Tor Inaba strain N16961 inoculum 10 days post-vaccination. A second group was challenged 3 months post-vaccination, while individuals in the third group remained unchallenged. A total of 197 volunteers participated in the trial. This dataset includes demographics and clinical and immunology measures from the 3 study sites: University of Maryland, Baltimore School of Medicine, Cincinnati Children's Hospital Medical Center, Ohio, and the Vaccine Testing Center, University of Vermont College of Medicine. (National clinical trial registry number: NCT01895855; Research funded by: PaxVax, Inc.)

## Timeframe

2013 - 2014

## Geographic Coverage

Baltimore (Md.)  
Burlington (Vt.)  
Cincinnati (Ohio)

## Subject of Study

[Human](#)

## Subject Domain

[Vaccines](#)

## Population Age

Adult (19 years to 64 years)

## Subject Gender

Female  
Male

## Keywords

[Cholera Vaccines](#)  
[CVD 103-HgR](#)

[Access via Application](#)

[Online Application Form](#)

## Access Restrictions

Application Required  
PaxVax Approval Required

## Access Instructions

[Submit Application](#)

## Associated Publications

[Chen WH, Cohen MB, Kirkpatrick, BD, Brady RC, Galloway D, Gurwith M, Hall RH, Kessler RA, Lock M, Haney D, Lyon CE, Simon JK, Szabo F, Levine MM. Single-dose live oral cholera vaccine CVD 103-HgR protects against human experimental infection with vibrio cholerae O1 El Tor. Clinical Infectious Diseases 2016; 62:1329-35](#)

## Data Type

Clinical Measures  
Demographic  
Immunology Laboratory Measures

## Equipment Used

Paper Source Documents

## Software Used

Proprietary Record Collection Database

## Study Type

Interventional

## Dataset Format(s)

Microsoft Excel

## Dataset Size

Variable Based on Data Requested

## Grant Support

Fully Funded by PaxVax, Inc./PaxVax, Inc.

## Access to Datasets Vary

- Contact author or assistant via email
- Application form
- Repository link where the researcher placed data

# Challenges

- University has no centralized, mandated repository for data
- Identifying research datasets
  - Publicly available – scattered
  - Hidden – local networks, personal devices
- Researchers don't have time to contribute or are concerned about sharing data
- Developing and maintaining software, esp. for interoperability

# Lessons Learned

- Team work is essential: library-wide and DCCP
- Need patience and perseverance
- Metadata librarian with science background is invaluable
- IT support is critical

# Accessing the UMB Data Catalog

- URL

<https://datacatalog.hshsl.umaryland.edu>

- HSHSL website or Google search

- Google Dataset Search

<https://toolbox.google.com/datasetsearch>

Presentation located at  
[https://archive.hshsl.umaryland.edu/handle/  
10713/8866](https://archive.hshsl.umaryland.edu/handle/10713/8866)

Thank you!