



My Favorite Health Data Resources

An overview and demonstration of useful health data tools and resources for researchers

CLINICAL AND TRANSLATIONAL SCIENCE CENTER

Christy Navarro, M.Sc.
CTSC Health Informatics Research Data Officer

Health Equity Resources and Outreach (HERO)
Program

Objectives

1. Locate health data tools and services available through Clinical & Translational Science Center
2. Wrangle disparate research tools, sites, and resources to generate a cohesive plan for identifying and organizing your research dataset
3. Identify and navigate Health Equity Resource and Outreach (HERO) program data resources.

UC Davis CTSC Health Informatics Research Data Officer

Underscoring the Importance of Comprehensive Support within the Health Equity Resources and Outreach (HERO) program

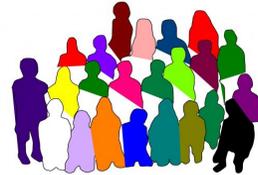
Introduction

In 2021 at the start of its 4th CTSA award, the UC Davis Clinical and Translational Science Center (CTSC) created the Health Equity Resources & Outreach (HERO) program. The addition of a Health Informatics (HI) Research Data Officer enables HERO to make data resources widely accessible and develop critical data fluency among researchers. The newly formed team is poised to support clinical researchers, public scientists, and community-based organizations seeking to make effective use of data to advance human health and promote inclusive science.



HERO Program

- Health Equity Tools & Resources
- Public Events & Series
- Networks
- Services
- Seed Grants



HI Research Data Officer

Essential Functions

- Education & Information Services
- HERO Informatics Program Development
- Outreach to Research Community

Alignment w/ Health Informatics AIMS to

- Expand access to research informatics, data management and expertise
- Integrate data science with informatics through training
- Develop new secure data standardization

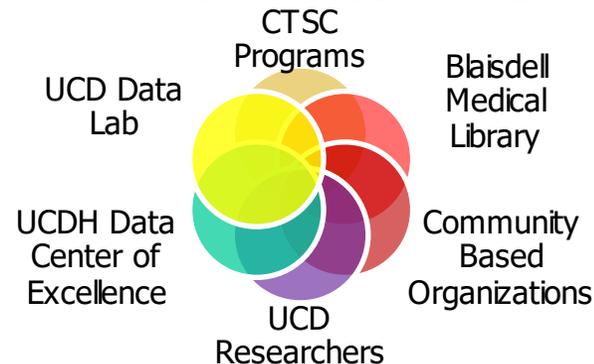


Biomedical Data Lifecycle
© 2021 by the President and Fellows of Harvard College
<https://datamanagement.hms.harvard.edu/about/what-research-data-management/biomedical-data-lifecycle>

What does data equity mean...?

“... It refers to the consideration, through an equity lens, of the ways in which data is collected, analyzed, interpreted, and distributed. It underscores marginalized communities' unequal opportunities to access data and, at times, their harm from data's misuse.”

Serving & Supporting



Early Projects

-  Subject guide in collaboration with Blaisdell Medical Library
-  Health Data Oversight - Data Sharing Committee
-  Health Data Equity Consultations
-  Education on data sharing plans and data management best practices
-  Environmental scan to identify interoperability and scalability of resources

Health Data Resources Environmental Scan

1. Interviews with Stakeholders
2. Data Fluency Poll



Feedback from Interview with Stakeholders

“We spend a ton of time talking to junior researchers about 101 data concepts”

“We need a community of data fluent clinicians and clinically fluent analysts and programmers”

Preliminary Highlights from Poll

- 75 respondents so far
- Data Collection
 - 63% interest in discovering data sources
 - 57% interest in appraising data quality
- Data Management
 - 58% interest in data organization/data manipulation
 - 57% interest in data conversion
 - 46% interest in metadata creation and use

Preliminary Highlights from Poll (cont.)

- Data Evaluation
 - 60% interest in data interpretation/data analysis (biostatistics)
 - 56% interest in data visualization
 - 50% interest in data tools
- Data Application
 - 54% interest in data sharing
 - 41% interest in data citations/data ethics

Observations

- Health Data Services:
 - Siloes
 - Often tailored for individuals who already know the process
 - Purpose and use of data influences priority (research vs. treatment or healthcare operations)
- Content, tools and resources are distributed across multiple sites and individuals

Poll #1 Future workshops

Which of these research topics do you want to learn more about?

- health research data lifecycle
- intro to visualizations
- data sharing/data management plans
- intro to metadata
- intro to data ethics, best practices and standards
- finding and utilizing open health data sets
- I would attend workshops or presentations on any of these topics

Considerations for inclusion:

- Utility
- Cost
- Ease of Learning
- Support for Users
- Relevance to the Research Data Lifecycle
- Tool or Resource Not Well Known

Biomedical Data Lifecycle © 2021 by the President and Fellows of Harvard College

<https://datamanagement.hms.harvard.edu/about/what-research-data-management/biomedical-data-lifecycle>



Favorite Health Data Tools for Researchers

1. Airtable
2. HERO Website Tools (open data resources)
3. Data Management Plan (DMP) Tool
4. Open Refine
5. Slicer Dicer
6. The Data Literacy Project (for fun)
7. UC Davis Profiles

Each researcher is responsible for understanding the classification of data they are working with so they can ensure the highest protections of the data. Not all of these are PHI approved.



What is Airtable?



Airtable

<https://www.airtable.com/product>

- Content, process, and project management tracking tool
- Easy to learn with an intuitive interface
- Cloud-based, making collaboration easy
- Option to generate embed code to display content on a webpage
- Various pricing models, including a free version
- Check out the health equity resources we created with Airtable
<https://health.ucdavis.edu/ctsc/area/HERO/health-equity-resources.html>
- Video of the demonstration for 2022 MLA My Favorite Tools
<https://bit.ly/AirtableMFT2022>

HERO Website Tools

- Health Equity Resources using Airtable
 - UC Davis Library Guide
 - Websites
 - Data Resources
 - UC Davis Data Assets (searchable list)
 - <https://health.ucdavis.edu/ctsc/area/HERO/health-equity-resources.html>
- Real World Health Data Catalog:
<https://airtable.com/shriVX6xK7nLTloGq>

Why We Need Data Management Plans



Build your Data Management Plan

Effective January 25, 2023, investigators are required to:

- **Submit** a Data Management and Sharing plan outlining how scientific data and any accompanying metadata will be managed and shared, considering any potential restrictions or limitations.
- **Comply** with the Data Management and Sharing plan approved by the funding Institute or Center (IC).
- **Previously** Data Management and Sharing Plans were only required for certain funders and amounts over \$500,000.



Build your Data Management Plan

- Data management plans are brief (2-3 page) documents that outline in advance how you will manage your data throughout the life of your project.
- The DMP Tool from the California Digital Library is an online tool for creating data management plans.
- The DMP tool uses templates to guide you through the process of creating a data management plan that meets funder requirements.

Why I like it

- ✓ Free
- ✓ Beginner friendly with templates and examples
- ✓ User resources through NIH and UC Davis Library for DMP reviews

Plan Overview

A Data Management Plan created using DMPTool

Title: Using natural language processing to determine predictors of healthy diet and physical activity behavior change in ovarian cancer survivors

Creator: Damian Yukio Romero Diaz -**ORCID:** [0000-0003-4661-0296](https://orcid.org/0000-0003-4661-0296)

Affiliation: University of Arizona (arizona.edu)

DMP ID: <https://doi.org/10.48321/D1BK5T>

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PAR-18-018

Grant: <https://reporter.nih.gov/search/qfhaBJoM20qq64VSqwCScg/project-details/10109452>

Template: DataWorks! Data Management and Sharing Plan Challenge

Project abstract:

Cancer survivors are a growing population in the United States; more than 16 million currently live in the US and by 2030 this number is expected to exceed 22 million. It is estimated that more than 50 percent of new cancer cases could be eliminated through a combination of healthy behaviors (e.g., physical activity and healthy diet); and cancer survivors are at high risk for developing new and recurrent cancer. Unfortunately, a significant percentage of cancer survivors are not attaining the cancer preventive guidelines of healthy diet and physical activity. In the past few decades, a variety of telephone-based lifestyle interventions have demonstrated effectiveness in helping survivors meet cancer preventive guidelines, however these trials are labor intensive and expensive to deliver, limiting their potential for broad dissemination. We propose to address this hurdle by taking advantage of recent advances in artificial intelligence to reduce the cost and maximize the impact of these much-needed interventions. Machine learning (ML) and Natural Language Processing (NLP) are analytical techniques that automatically learn from direct and indirect patterns in data. We propose to use machine learned algorithms to analyze speech to aid in predicting who may be at risk of poor adoption of healthy lifestyle behaviors. These speech data will come from the Lifestyle Intervention for Ovarian cancer Enhanced Survival (LIVES) study, a telephone-based lifestyle intervention testing whether a diet low in fat and high in vegetables, fruit, and fiber, coupled with increased physical activity will increase time to disease progression in 1200 ovarian cancer survivors who have recently completed treatment, as compared to an attention control. Intervention coaches employed motivational interviewing to elicit behavior change and all calls on the LIVES trial were recorded with repeat assessments of diet, physical activity, patient reported and clinical outcomes. We will use this existing and robust longitudinal data set, which pairs conversational speech data with explicit outcomes, to achieve the following objectives. 1) Develop a ML model to identify patterns in the interactions between coaches and their participants that signal a likelihood of optimal behavior change in diet and physical activity given the comprehensive LIVES data set, utilizing voice recorded calls, demographics, and clinical and patient reported outcomes collected at multiple time points. 2) Decompose the ML model in terms of “intervenable factors”, so that participant affect, coach adherence to the intervention protocol, and other important aspects of the interaction can be individually evaluated for their role in predicting behavior change, as well as adherence to intervention goals. This decomposition will directly enable early and targeted adjustments to intervention plans for individuals, reducing the cost and increasing the efficacy of intervention strategies. ML and NLP methods can produce models that listen to a coaching conversation and automatically predict whether it will result in positive change towards enactment of healthy lifestyle behaviors. Such predictive models would enable more efficient, effective, and individualized lifestyle interventions, the first step towards personalized behavioral medicine.

Start date: 01-01-2021

End date: 12-31-2022

Last modified: 02-20-2022

What is OpenRefine?



www.openrefine.org

Open Refine (formerly Google Refine) is a free, open-source tool for cleaning data. It runs locally in your browser.

Why I like it

- ✓ Free
- ✓ Beginner friendly
- ✓ No coding required
- ✓ User learning resources

Open Refine – Demonstrations



Demonstration - Introduction (4:13)

Other demonstrations to check out:

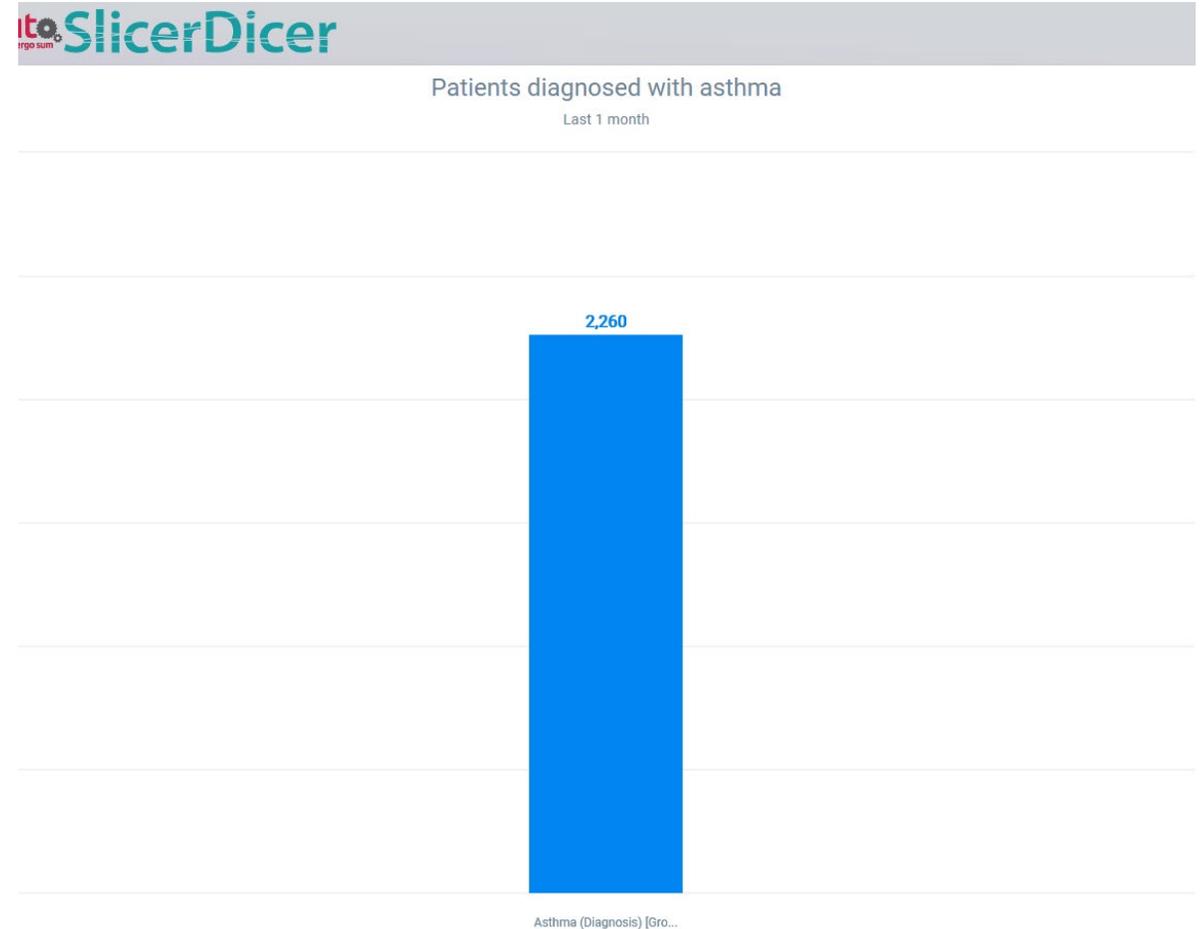
- Data Transformation (8:33)
- Data Augmentation (6:10)

Cleaning, Reconciling with Wikidata

<https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing/Tutorials/Video>

SlicerDicer – what is it?

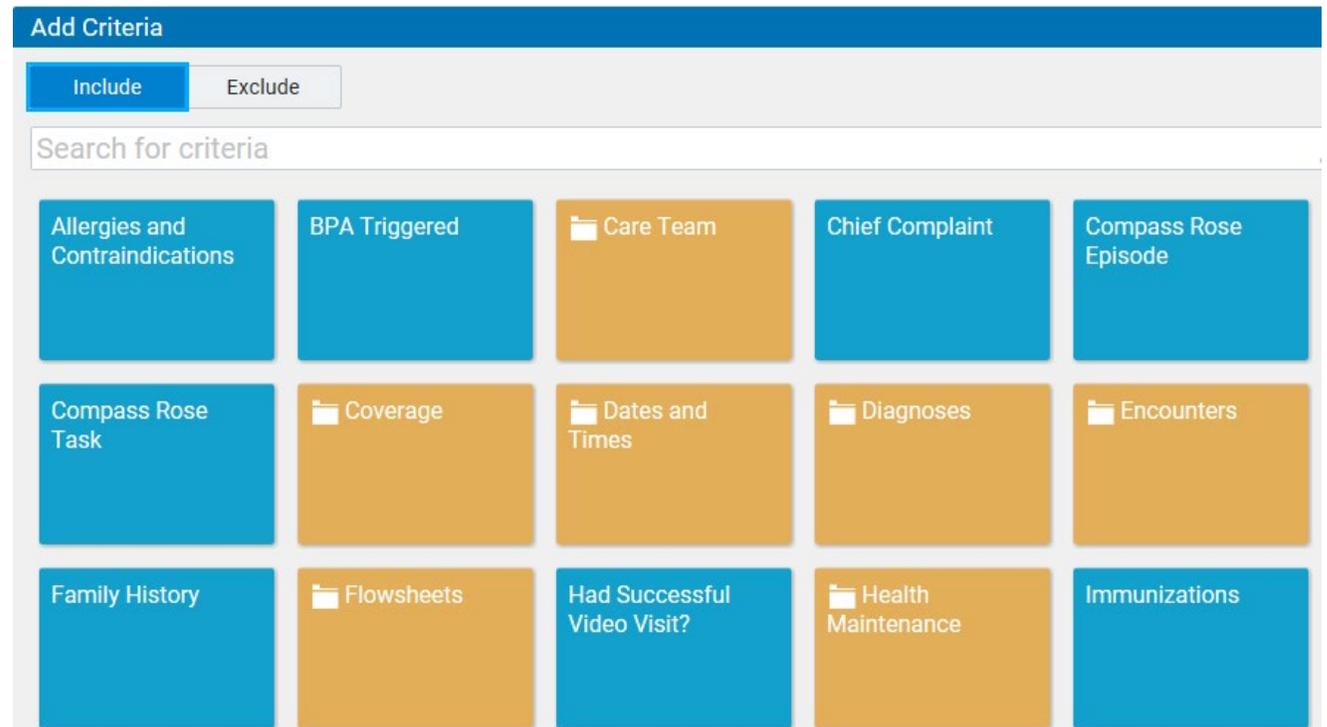
- Self-service reporting tool
- Fast, easy data exploration
- Provides a graphical response
- Identify patient populations for research cohorts (de-identified data only)



SlicerDicer Value – Uses

- Use combinations of inclusion/exclusion criteria to identify specific patient populations:

- Examples:
 - Patient demographics
 - Immunizations
 - Procedures
 - Labs
 - Medical history
 - Surgical history
 - Etc...



SlicerDicer (cont.)

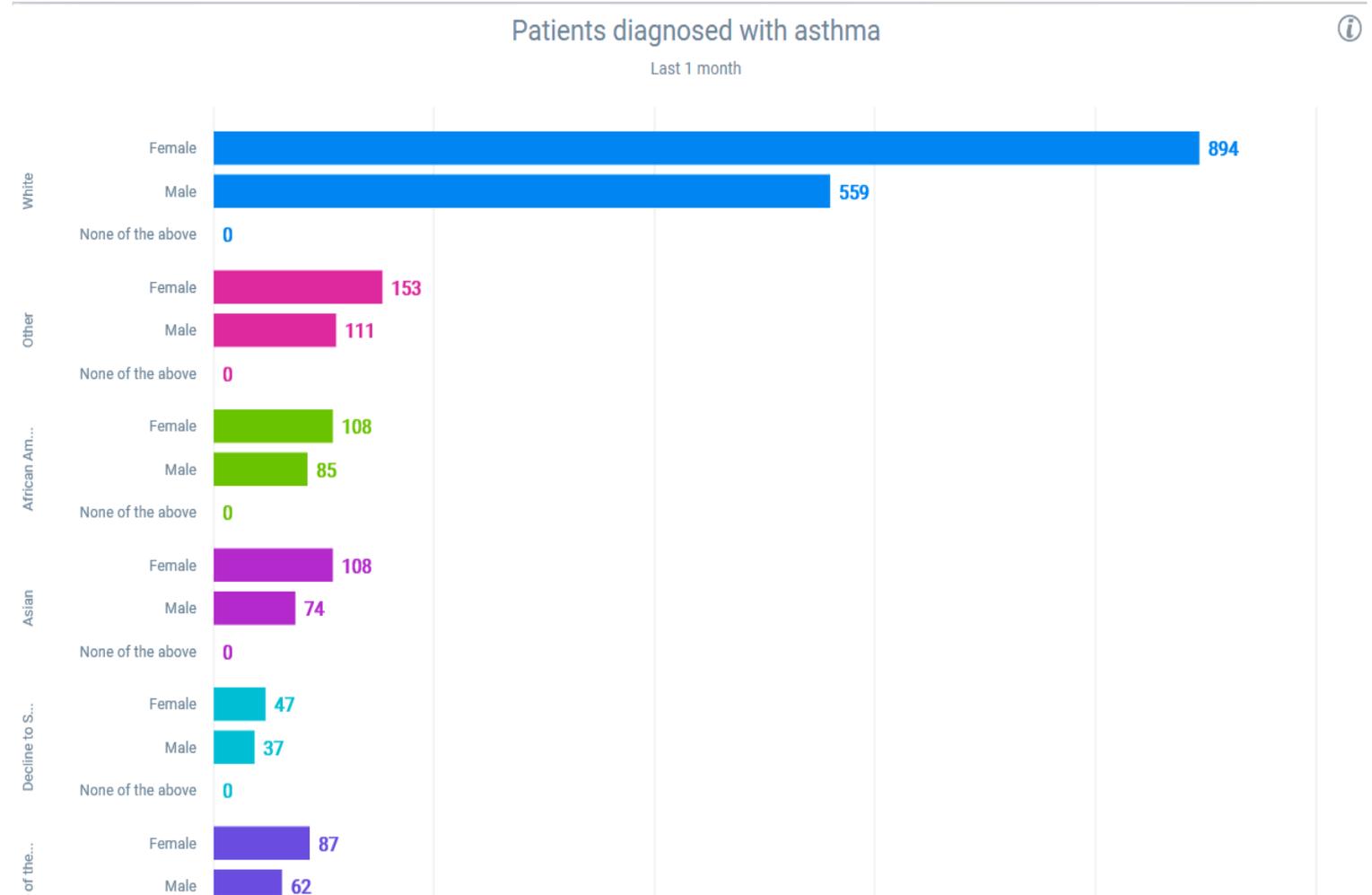
Slice populations to drill down even further

Example:

You've identified patients who have been diagnosed with asthma in the past month.

Slice by:

- Gender
- Race



Slicer Dicer Resources

The screenshot shows the UC Davis Health website's 'Epic and IT Education' section. The page features a search bar at the top right and a navigation menu on the left. The main content area is titled 'Cogito SlicerDicer' and includes a description of the tool, a list of training guides, tip sheets, and request training links. The navigation menu on the left includes categories like 'Epic Quarterly Updates', 'Epic Basics', 'Upcoming/Current Projects', 'Current IT Initiatives', and 'Patient Movement'.

UC DAVIS HEALTH | Epic and IT Education

Search all Search this site

EMR News Operational Resources Contacts

Epic Quarterly Updates
Quarter 3 - 2022 Release
Quarter 2 - 2022 Release
Quarter 1 - 2022 Release
Quarter 4 - 2021 Release

Epic Basics

Upcoming/Current Projects
Ambulatory MAR
COVID-19 Epic EMR Updates & Enhancements
CPM 2022 Upgrade
Decision Tree
Direct Admits
E-Consults and Smart Referrals ePA
ERAS (Enhanced Recovery After Surgery)
Haiku/Canto Mobile Epic Access
Hello Patient
Kaleidoscope
Lumens
MyUCDavisHealth (MyChart)
MyChart Message Reduction (MMR)
Pain Management Updates
Patient Estimates
Patient Questionnaires
Phoenix Transplant Module
Results Review
Secure Chat
Specialty Pharmacy - Compass
Rose
Storyboard
Surge Expansion
Symptom Checker/Self-Triage
Welcome

Current IT Initiatives
EPCS MFA Change
EPCS-DUO

Patient Movement

How To....
Request a Change to Workqueue
Ownership
Request for IT

Cogito SlicerDicer

SlicerDicer is a self-service reporting tool that provides physicians, department managers, and other users with intuitive and customizable data exploration abilities. Using SlicerDicer, users can find the data they need to investigate a hunch, and then refine their searches on the fly to better understand the data they work with. Right in Hyperspace, they can examine trends, drill down to line-level details, and jump to related records to follow up.

With SlicerDicer, users can search for free text in lab results, combine similar elements on the fly, and jump into patient charts. If you're not sure where to start, enter a keyword and select from a list of suggested search criteria. Then, choose a visualization that matches your needs, including vertical and horizontal bar graphs, line graphs, maps, and tree map charts. SlicerDicer can show results using a variety of different measures, including totals, percentages, averages, variance, maximums, and minimums. For more details see the training materials below.

Training Guides

- SlicerDicer Quick Start Training Guide
- SlicerDicer Advanced Training Guide
- PowerPoint Overview

Tip Sheets

- How to Limit Results by Service Area/Location/Department
- SlicerDicer Filter for Research
- Research Request for PHI

Request Training/Access/Help

- To request SlicerDicer basic training, click here
- To request SlicerDicer Permission, Data Model Request, Custom Filter Request and Data Questions, click here

Virtual Office Hours

Virtual office hours are available via WebEx and are hosted by members of the SlicerDicer team.

Click a link listed below to access the WebEx meeting on the date and time you want to attend.

New office hours:

Occurs every Monday from 12:00 PM to 1:00 PM
<https://ucdhs.webex.com/ucdhs/j.php?MTID=m435d44558a3fd1ee4a7b9ef928fbde8>

Occurs every Wednesday from 9:00 AM to 10:00 AM
<https://ucdhs.webex.com/ucdhs/j.php?MTID=m28be55ab8f0e88ddd88a0e2afa54a9ec>

Occurs every Thursday from 1:00PM to 2:00PM
<https://ucdhs.webex.com/ucdhs/j.php?MTID=m4deb30e92a25751cad7dc7208112d019>

FAQs

Coming Soon!

EMR Research team:

EMRResearch@groups.ucdavis.edu

For additional SlicerDicer information and training materials, visit:

[Slicer Dicer \(ucdavis.edu\)](http://ucdavis.edu)

[SlicerDicer - UC Davis Clinical Research Guidebook - Confluence](#)

For EMR Access:

[EMR \(Electronic Medical Record\) Access New/Change - IT Self Service \(service-now.com\)](#)

For SlicerDicer Access:

[Epic SlicerDicer Request - IT Self Service \(service-now.com\)](#)

UC Davis Profiles profiles.ucdavis.edu

The screenshot shows the top navigation bar with the text 'es' on a dark blue background. Below it is a search bar with 'History (0)', 'Search Options', and a dropdown menu set to 'Everything'. A link '(photo, awards, links to other websites, etc.)' is visible. Two buttons, 'Find People' and 'Find Everything', are present. The main heading is 'Find People by Research Topic or Name'. The search form includes a 'Research Topics' input field with a 'Search' button, followed by 'and/or'. Below are input fields for 'Last Name', 'First Name', 'School', 'Department', 'Researcher Type', and 'More Options', each with a dropdown arrow. A 'Student Projects' checkbox is at the bottom.

What is it?

- A research networking and expertise discovery platform that represents a large network of institutions

How its useful to researchers?

- Find people
- Discover networks
- Learn about experts

Why I like it

- Free
- Open to the public
- Easy to use
- Includes Profile information for: UC Davis, UC Irvine, UCLA, UCSD, UCSF & USC

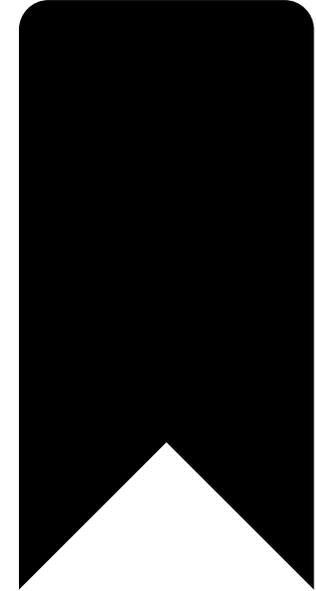
Poll # 2 Most Useful

Which of the tools is most useful in your work? Why?

- Airtable
- HERO Website Tools
- DMP Tool
- Open Refine
- UC Davis Profiles

Link to a full list of My Favorite Tools

- <https://airtable.com/shrvmZJspP7JSYKzt>



Poll # 3 Future Workshops

Which of the following future workshops would you attend?

- Airtable
- HERO Website Tools
- DMP Tool
- Open Refine
- Slicer Dicer
- UC Davis Profiles

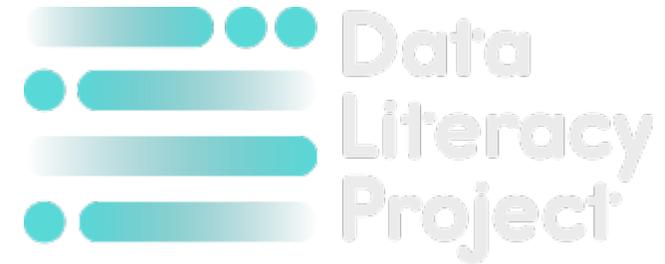
5 Minute Activity: Put your data literacy to the test!

<https://thedataliteracyproject.org/assessment>

What I like

Provides a jumping off point for your next Data Fluency Upgrade* (book recommendations etc.)

Free and open to anyone



Future Events

Slicer Dicer Workshop

October 4, 2022

Jenny Neeley

Spotlight # 3– Categorization of Human Populations in Biomedical Research

October 31, 2022

Alice Beecher Popejoy, Ph.D.

For notifications of Future Events Join the CTSC Events
Calendar:

<https://health.ucdavis.edu/ctsc/event-calendar.html>



A person wearing a blue lab coat is holding a white rectangular sign with both hands. The sign has the word "QUESTIONS?" written in large, bold, dark blue capital letters. The background is a blurred grey.

QUESTIONS?

Contact



Christy Navarro

cenavarro@ucdavis.edu

<https://health.ucdavis.edu/ctsc/area/HERO/health-equity-resources.html>

Health Data Resources for UC Davis Health Researchers

- Biomedical Informatics (CTSC)
- Biorepositories (CTSC)
- Blaisdell Medical Library
- Health Equity Resources and Outreach
- (CTSC)
- Clinical Research Ethics (CTSC)
- Clinical Trials Office (CTSC)
- Community Engagement (CTSC)
- Data Center of Excellence
- Research Training (CTSC)
- School of Medicine Office of Research
- UC Davis Data Lab
- UC Davis IRB

Facilitation Links

Slide #9 Run Poll #1

Slide #(just in case) Slide 12 Demo Video YouTube

<https://bit.ly/AirtableMFT2022>

Slide 13 – Health Equity Resources Demonstration

<https://health.ucdavis.edu/ctsc/area/HERO/health-equity-resources.html>

Slide 13 – Real World Health Data Catalog

<https://airtable.com/shriVX6xK7nLTloGq>

Slide 19 <https://thedata literacyproject.org/assessment>

Slide 22 Poll #2 Most Useful

Slide 23 Poll # 3 Future Workshops