Science collaborative for Health disparities and Artificial intelligence bias Reduction

Deborah Duran, PhD
Luca Calzoni, MD MS
NIMHD

National Institute on Minority Health and Health Disparities

National Institute of Nursing Research

Office of Data Science Strategy

ScHARe is a **cloud-based platform with population science**, including **social determinants of health**, data sets designed to accelerate research in health disparities, health and healthcare delivery outcomes, and artificial intelligence (AI) bias mitigation strategies.

ScHARe aims to fill three critical gaps:

- Increase participation of **women & underrepresented populations with health disparities** in data science through data science skills training, cross-discipline mentoring, and multi-career level collaborating on research.
- Leverage population science, SDoH, and behavioral Big Data and cloud computing tools to foster a **paradigm shift** in health disparity, and health and healthcare delivery outcomes research.
- **Advance AI bias mitigation and ethical inquiry** by developing innovative strategies and securing diverse perspectives

nimhd.nih.gov/schare

ScHARe

NINR encourages researchers to view the **health equity** and **social determinants of health research** lenses as the primary foci through which to consider other research lenses

ScHARe is designed as a cloud platform for Health Equity and Social Determinants of Health data science and AI Bias mitigation research

NINR's Strategic Plan

Phase I

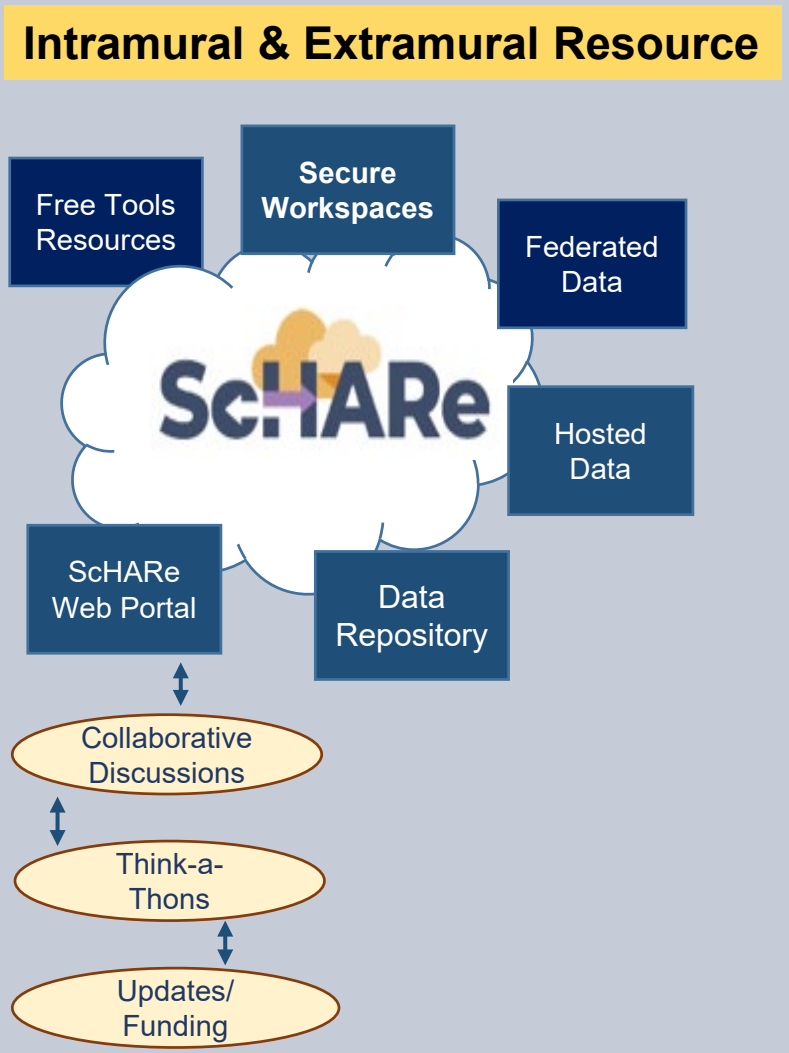
Population Science and SDoH Datasets
Tutorials and Resources
Think-a-Thons

ScHARe Components

ScHARe co-localizes within the cloud:

- **Datasets** (including social determinants of health and social science data) relevant to minority health, health disparities, and health care outcomes research
- **Data repository** to comply with the required hosting, managing, and sharing of data from NIMHD- and NINR-funded research programs
- **Computational capabilities and secure, collaborative workspaces** for students and all career level researchers
- **Tools for collaboratively evaluating and mitigating biases** associated with datasets and algorithms utilized to inform healthcare and policy decisions

Frameworks: Google Platform, Terra Interface, GitHub, NIMHD Web ScHARe Portal



ScHARe Data Ecosystem

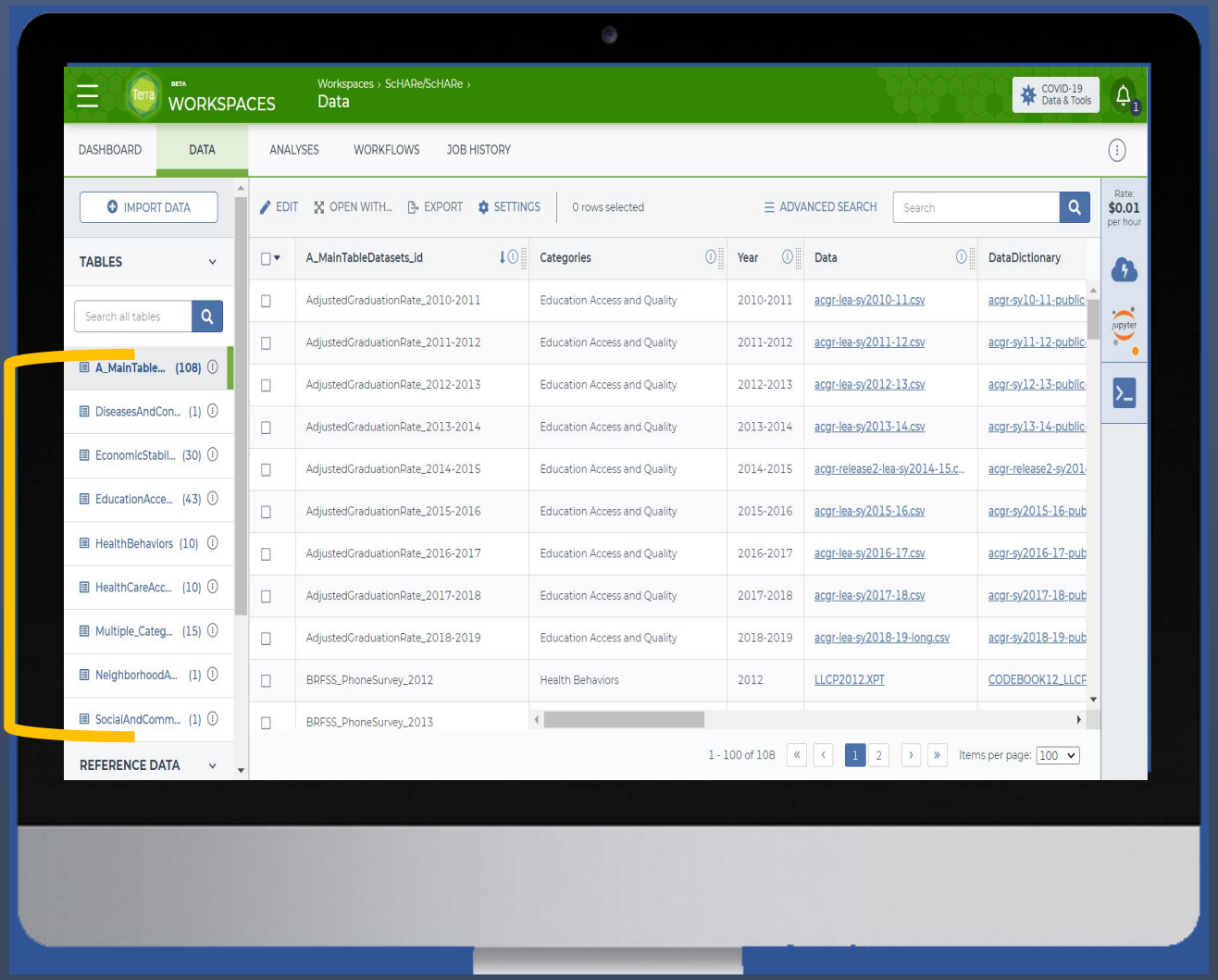
Researchers can access, link, analyze, and export **a wealth of datasets** within and across platforms relevant to research about health disparities, health care outcomes and bias mitigation, including:

- **Google Cloud Public Datasets:** publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program.
  Example: *American Community Survey (ACS)*

- **ScHARe Hosted Public Datasets:** publicly accessible, de-identified datasets hosted by ScHARe.
  Example: *Behavioral Risk Factor Surveillance System (BRFSS)*

- **Funded Datasets on ScHARe:** publicly accessible and controlled-access, funded program/project datasets using Core Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy.
  Examples: *Jackson Heart Study (JHS); Extramural Grant Data; Intramural Project Data*

Data organization



ScHARe

On ScHARe, datasets are categorized by content based on the CDC **Social Determinants of Health categories**:

1. Economic Stability
2. Education Access and Quality
3. Health Care Access and Quality
4. Neighborhood and Built Environment
5. Social and Community Context

with the addition of:

- **Health Behaviors**
- **Diseases and Conditions**

Users will be able to **map and link** across datasets

Data Overview

ScHARe Data Ecosystem will offer access to **300+ datasets of population science, SDoH, behaviors** including:

- Google Cloud Public Datasets
- ScHARe Hosted Public Datasets

In Discussion

- EPIC pulse oximeter data
- OMH demonstration data
- Jackson Heart Study
- Geospatial Aggregated

Coming in Fall: Repository for Funded Datasets on ScHARe – Compliance with NIH Data Sharing Policy

Data Examples

American Community Survey
Behavioral Risk Factor Surveillance System (BRFSS)
Youth Risk Behavior Surveillance System (YRBSS)
US Census Data
Area Deprivation Index (SES of communities)
Social Determinants of Health Database (AHRQ)
Community Resilience Estimates
Minority Health Social Vulnerability Index
Current Population Survey -Annual Social & Economic(Labor)
Current Population Survey-Food Security (USDA)
Food Access Research Atlas
National Health and Nutrition
National Household Education Survey Program
Dept of Education (EDFACTS DATA)
Medical Expenditure Panel Survey (MEPS)
National Environmental Public Health Tracking Network
Local Area Transportation Characteristics for Households (LATCH)
US Chronic Disease Indicators (CDI)
Mental Health Services Survey (NMHSS)
National Survey of Substance Abuse Treatment Services (NSSATS)
FBI Hate Crimes Statistics

Phase II (in process)

Repository and Data Ecosystem

ScHARe Data Repository

CORE COMMON DATA ELEMENTS

COMPLY WITH DATA SHARING POLICY - HOST PROJECT DATA

NOVEL CDE FOCUSED REPOSITORY TO FOSTER INTEROPERABILITY

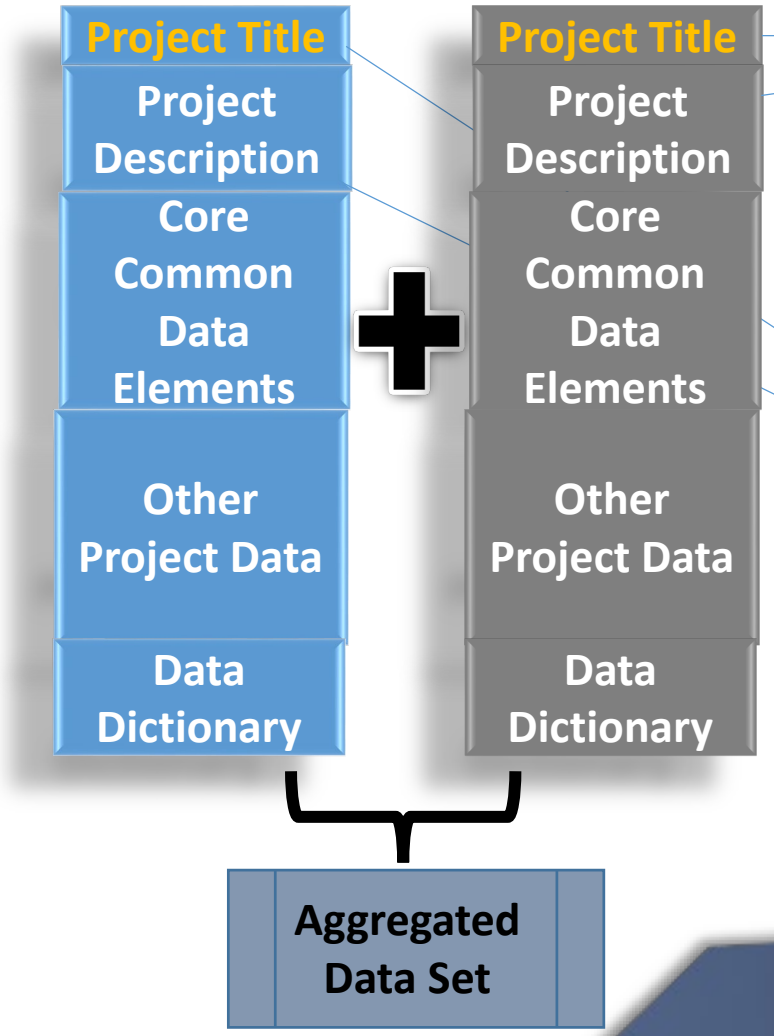
DATA ECOSYSTEM

- Map across datasets
- Map across platforms

UPCOMING

Core Common Data Elements
Intramural and Extramural Project Repository

- Compliance with NIH Data Sharing Policy
- Fosters dataset sharing and interoperability by using or mapping to core common data elements
- Provides resources for intramural researchers to work in a secure workspace and host data
- Centralized aggregated data sets for repeat use
- Maps across project data



UPCOMING FALL

Project & Federated Data Set Mapping

- Project Title
- Project Description
- Core Common Data Elements
- Other Project Data
- Data Dictionary

\+ AMERICAN COMMUNITY SURVEY

\+ Medical Expenditure Survey (MEPS)

\+ Pharmacy and health insurance databases

Mapping Across Cloud Platforms

UPCOMING

ScHARe Think-a-Thons (TaT)

- Monthly sessions (2 hours)
- Instructional/interactive
- Designed for new and experienced users
- Research & analytic teams to:
  - Conduct health disparities, health outcomes, bias mitigation research
  - Analyze/create tools for bias mitigation
- Publications from team collaboration
- Networking
- Mentoring and coaching
- Focus:

✓ **Instructional**
✓ **Collaboration Research Teams**
✓ **Bias Mitigation**
✓ **Target underrepresented pops events**

ScHARe

Think-a-Thon

Artificial Intelligence and Cloud Computing Basics

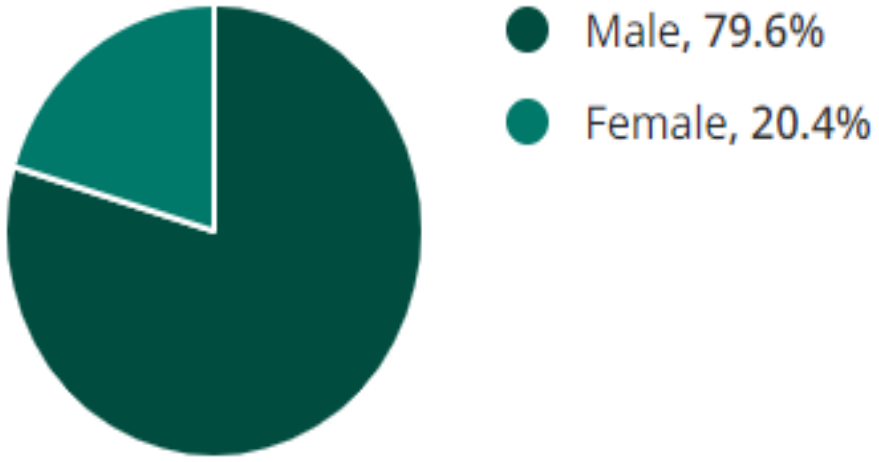
Terra: Data Sets and Analytics

Register:

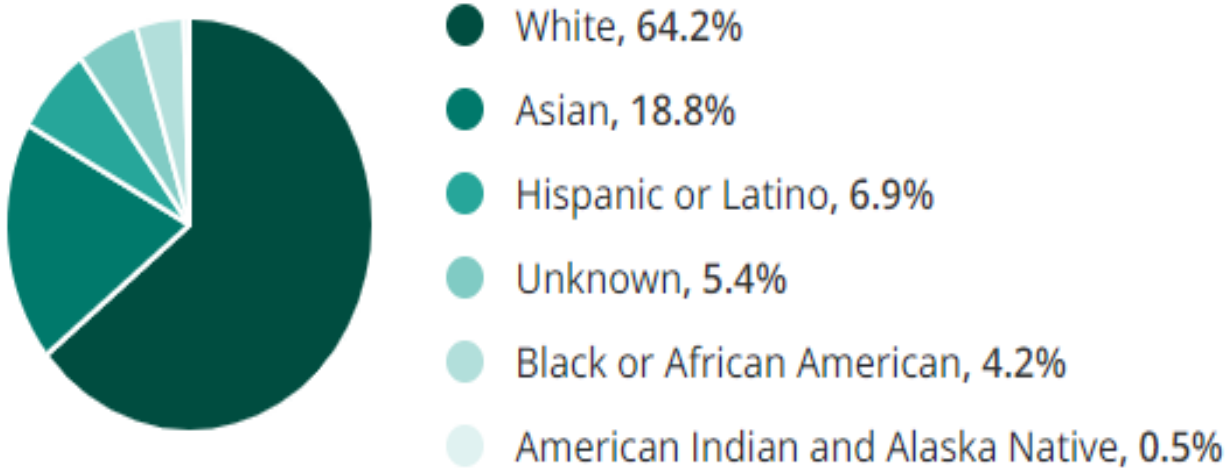
bit.ly/think-a-thons

Target Women and Health Disparity Populations Underrepresented in Data Science

DATA SCIENTIST STATISTICS BY GENDER



DATA SCIENTIST STATISTICS BY RACE



- ✓ Target low-resource institutions, including minority serving and community colleges
- ✓ Special Event Think-a-Thons targeting a gap area – ScHARe as a resource for Educators
- ✓ No cost tutorials on how to use Terra, which gives access to most NIH platforms
- ✓ $300 free credits
- ✓ Centralized population science, including SDoH, data sets
- ✓ Mentoring model for research projects that crosses the career pipeline back-n-forth

AI Bias Mitigation Strategies

- Widespread use of AI raises a number of ethical, moral, and legal issues – likely not to go away
- Algorithms often are “black boxes”
- **Biases can result from:**
  - **social/cultural context not considered**
  - **design limitations**
  - **data missingness and quality problems**
  - **algorithm development and model training**
  - **implementation**
- If not rectified, biases may result in decisions that lead to discrimination, unequitable healthcare, and/or health disparities
- Lack of diverse perspectives - Populations with health disparities are underrepresented in data science
- Guidelines and recommendations emerging from HHS, NIST, White House, etc.

ScHARe

Critical thinking can mitigate AI biases.

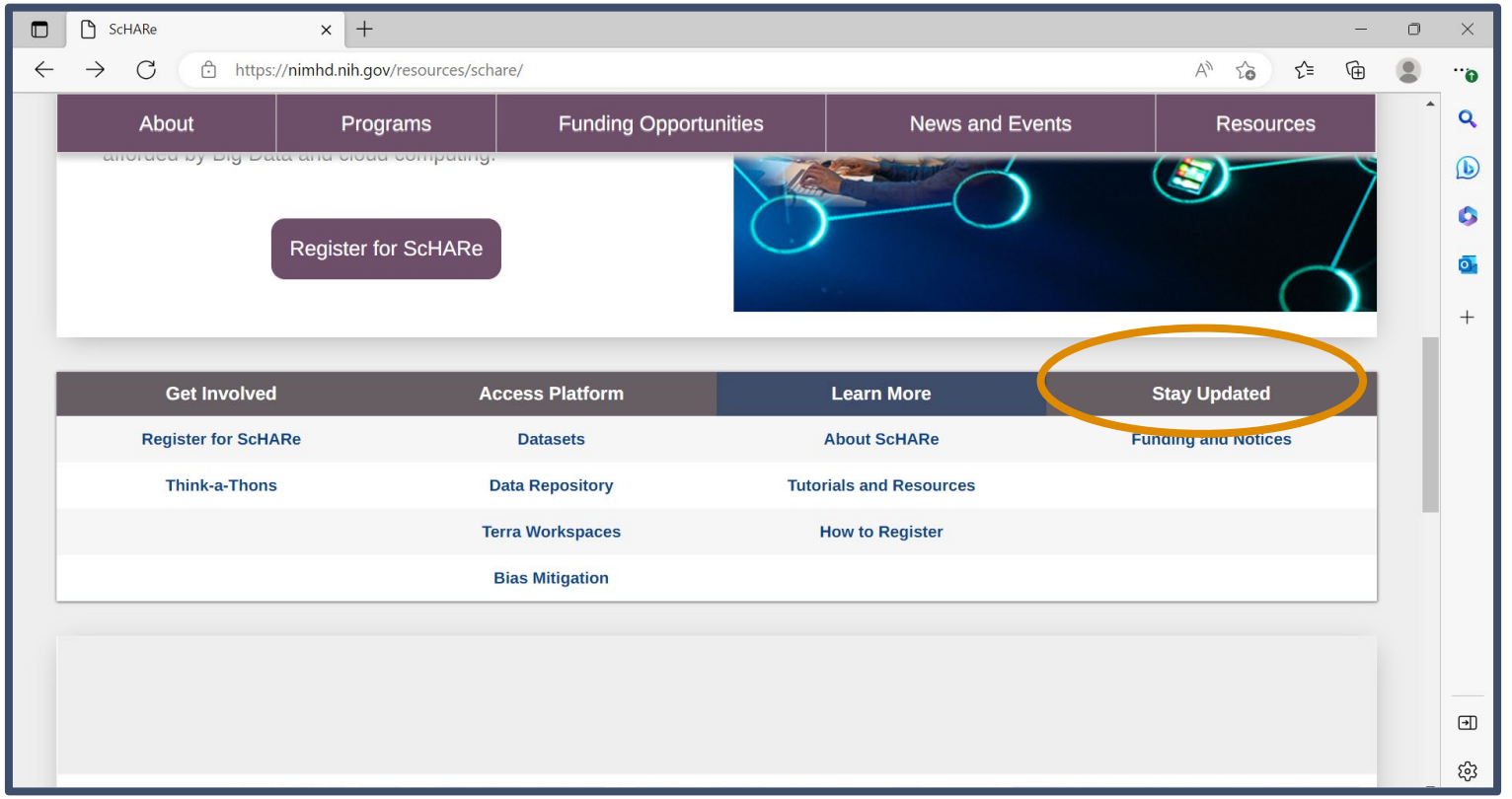
ScHARe was created to:

- foster participation of **populations with health disparities in data science**
- promote the collaborative identification of **bias mitigation strategies across the continuum**
- create a **culture of ethical inquiry** and critical thinking whenever AI is utilized
- build **community confidence** in implementation approaches
- focus on **implementation of AI bias** guidelines and recommendations

Coming Soon: “Bias Buster” Challenge

Two Ways to Sign up for ScHARe News

Scannable from your screen!



nimhd.nih.gov/schare

Next Think-a-Thons:

bit.ly/think-a-thons

Register for ScHARe:

bit.ly/join-schare

schare@mail.nih.gov

Thank you

NIMHD

Dr. Eliseo Perez-Stable

ODSS

Dr. Susan Gregurick

NIH/OD

Dr. Larry Tabak

NINR
Dr. Shannon Zenk

NINR
(Rebecca Hawes
Michael Steele and
John Grason)

***ORWH**

***OMH**

NIMHD OCPL
Kelli Carrington
Thoko Kachipande
Corrine Baker

BioTeam

STRIDES

Terra

SIDEM

RLA

CCDE Working Group

Deborah Duran
Luca Calzoni
Rebecca Hawes
Michael Steele
Kelvin Choi
Paula Strassle
Michele Doose
Deborah Linares
Crystal Barksdale
Gneisha Dinwiddie
Jennifer Alvidrez
Matthew McAuliffe
Carolina Mendoza-Puccini
Simrann Sidhu
Tu Le