

# Outcomes from Transitions in Education and Work: New Opportunities to Learn from Data

Julia Lane, New York University and the Coleridge Initiative

# Key ideas

- Leverage new opportunity to use data for evidence based policy
- Build on federal and philanthropic investments
- Apply technical and human approach to create high value products
- Partner in cross-state initiative

# Context

Action 7: Pilot an Automated Inventory Tool for Data.gov

Action 8: Pilot Standard Data Catalogs for Data.gov

Action 9: Improve Data Resources for AI Research and Development

## DRAFT 2019-2020 Federal Data Strategy Action Plan

Co-created by Federal agencies and the public, coordinated, and edited by the *Cross-agency Priority Goal: Leveraging Data as a Strategic Asset* Federal Data Strategy development team



Action 13: Assess Data and Related Infrastructure Maturity

Action 15: Identify Data Needs to Answer Key Agency Questions

Action 16: Identify Priority Datasets for Agency Open Data Plans

H.R. 1831: Evidence  
2016

Introduced: Apr 16, 2016  
114<sup>th</sup> Congress

Status: Enacted —  
This bill was

Law: Pub.L. 114-1

Sponsor:  R

Text:  R

ce-Based  
capacity for producing

evidence  
protections,  
d be used to  
unless

lications to  
searchers for  
ing a common  
cess restricted,

Coordination.  
y establishing a  
g coordination

**Capacity**

Enables  
requiring that  
data needs,

improves agency  
n by establishing  
requiring

sources  
ment agencies  
ties to engage  
ivities and use  
ment

are relevant to the needs businesses, policymakers, and the public. \$10.1 million increase

array of public use data.



Applied Data Analytics

Sponsored by:



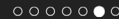
OFFICE OF FAMILY ASSISTANCE  
An Office of the Administration for Children & Families

# Context

## UPCOMING TRAINING PROGRAMS

### NCSSES - Alexandria, VA - Fall 2019

In this 10-day Applied Data Analytics training program, participants will work in teams to define and complete a project related to career pathways for doctoral recipients. The program will provide up-to-date perspective and hands-on instruction on using micro data in SQL and Python for tasks such as data management, record linkage, data visualization, and machine learning. The additional four days beyond the base 6 day program will provide additional project time and further analysis modules such as Network Analysis, text analysis and the use of APIs.



### Join waitlist

Application: **Closed**

Introduction to SQL & Python begins **November 4 (online)**

In-person training in Alexandria, VA

- Module 1: **December 4-6, 9, 10**
- Module 2: **January 15-17, 21, 22**

Remote Presentations: **TBD**

COLERIDGE  
INITIATIVE

USDA  
United States  
Department of  
Agriculture

PRESENT

## DATA ANALYTICS TRAINING:

the Food Purchasing Patterns of  
WIC Participants

## Outcomes from Transitions in Education and Work: Developing a Scalable Regional Approach

### Summary

There is enormous interest in building a better understanding of how people transition across different educational and work experiences to sustainable jobs. Transitions of people as they age can be nonlinear and include transitions within and across secondary and post secondary institutions, as well as the use of government services such as disability services, criminal justice interventions, or workforce development. Individuals can also move across political jurisdictions such as state or county lines, making it incredibly difficult to understand the regular patterns in service utilization. The sheer variation on types of transitions individuals can make and the geographical movement across borders opens up opportunities for governments to intervene in more productive ways. The data sharing across states and agencies also provides increasing possibilities for more efficient and effective utilization of government services among vulnerable populations.

Building a common framework across states requires (i) identifying key issues of interest (ii) bringing together data from many sources across both state and agency lines and (iii) bringing together staff from those agencies. These three activities are necessary to work through the important data and measurement issues that are necessary to build operational metrics for decision-making. This workshop will build on a successful collaborative approach developed by the State of Illinois, New York University's Coleridge Initiative, Chapin Hall and Ohio State University that uses an inter-agency and inter-state collaborative, classroom environment to address the issues. It will bring together key stakeholders from the MidWest to develop an operational agenda for the spring of 2019. We expect to discuss how to develop such metrics as:

1. What are the main educational and work flows of interest for policy?
2. What data are available?
3. What kind of metrics would be of greatest use to stakeholders?

# Approach: Technical and Human

## Technical

- Create secure environment where data providers can share their data across agency and jurisdictional lines
- Census Authorization to Operate; HHS and USDA in process

## Operational

- Link disparate data
- Analyze data

## Legal & Practical

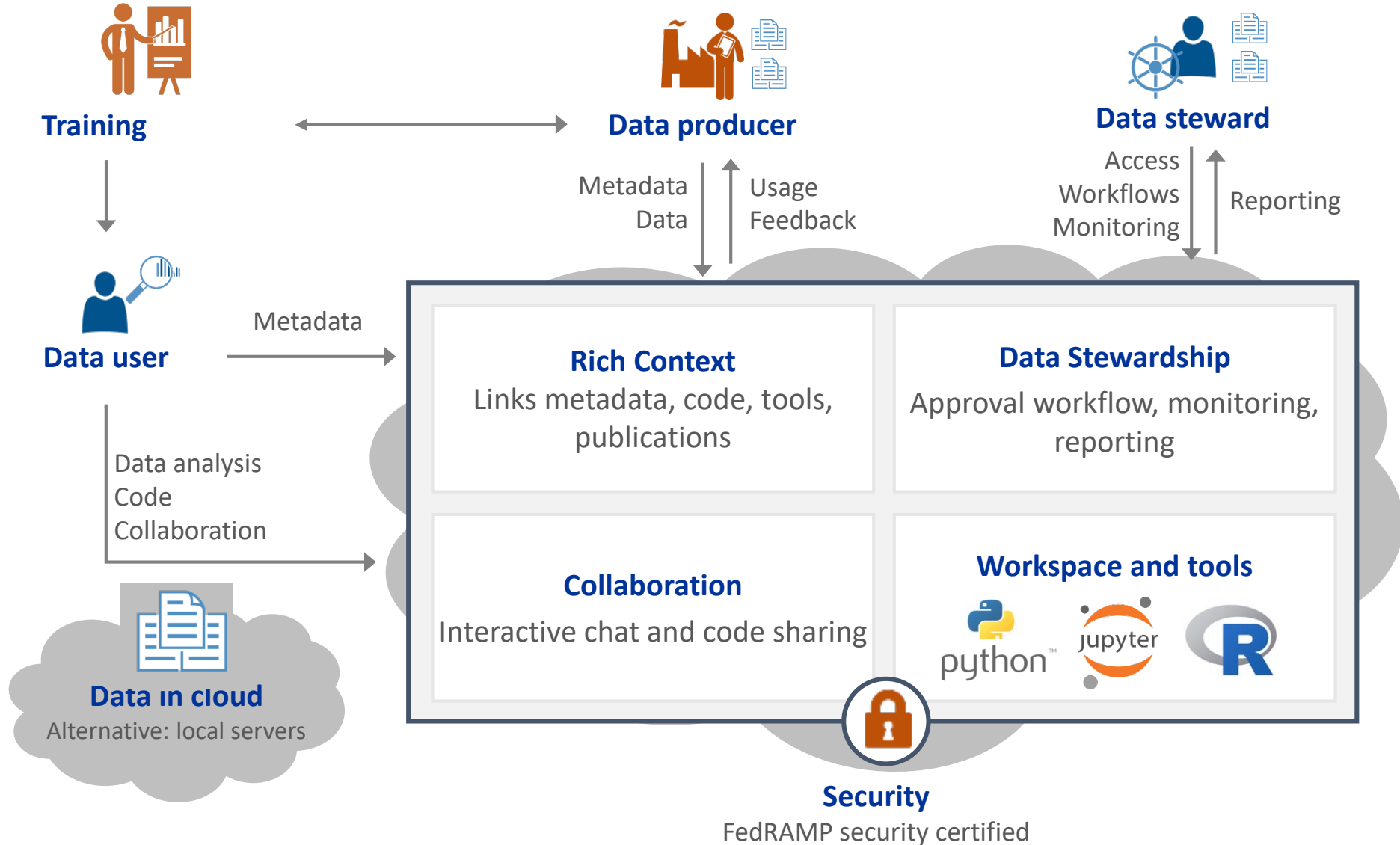
- Document value associated with the data linkage
  - Consistent with the agency mission
  - Useful enough to engage decision-makers

# Technical Approach: Safe Use

- Safe Projects
  - Approved projects, consistent with agency mission
- Safe People
  - Approved and trained researchers
- Safe Settings
  - Secure environment, GovCloud, FedRamp Moderate
- Safe Data
  - Deidentified Data
- Safe Outputs
  - Disclosure reviews and export controls

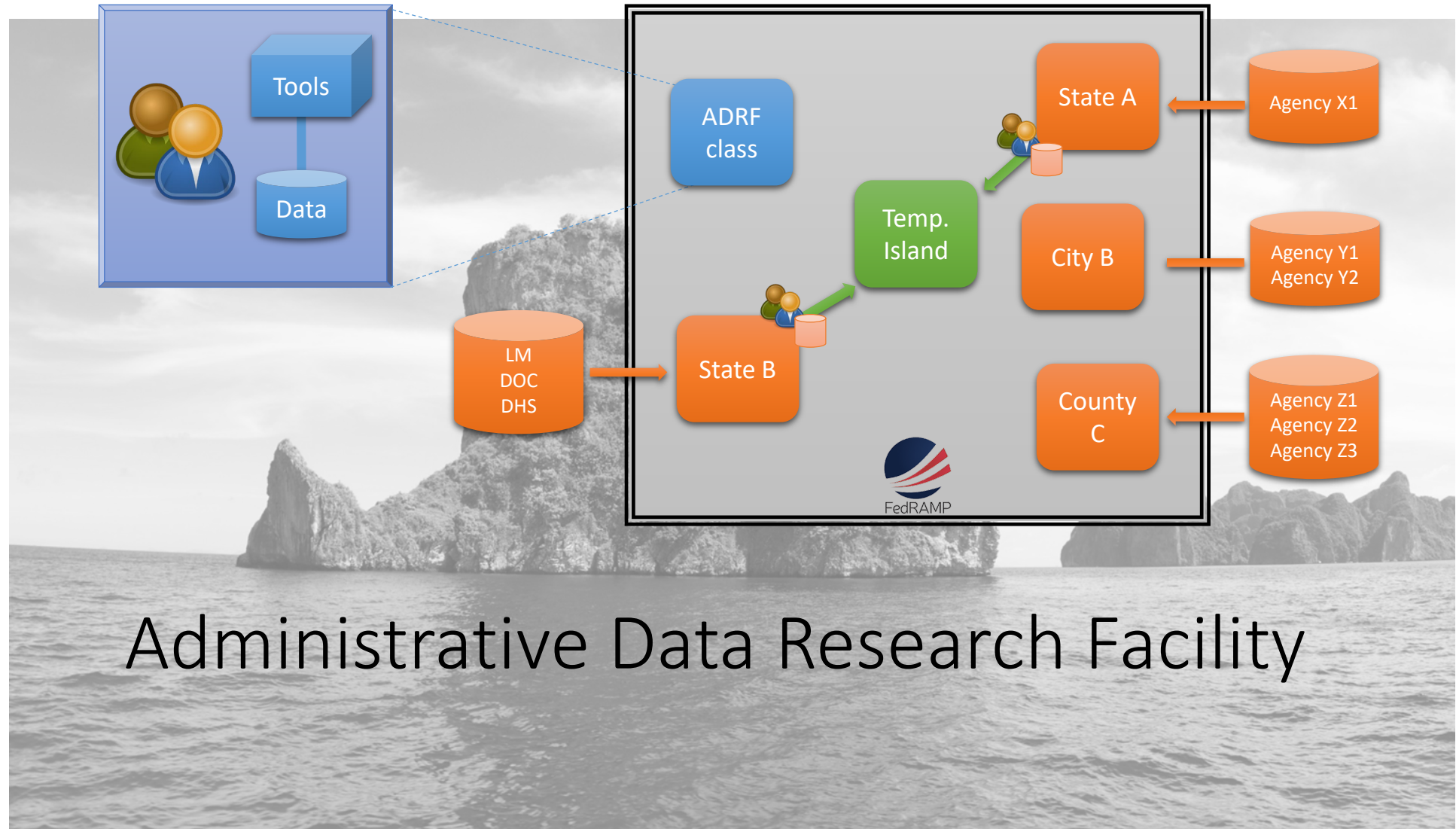
= SAFE USE

# Technical: Operational Environment





# Technical: Access



## Administrative Data Research Facility



# Technical: Automate Data Approval Workflow

ADRF | Data Stewardship

Navigation

Projects

Explore Data

User Directory

Bookmarks

Project Request

Please provide the following information to initiate a project request in the ADRF. Your request will be automatically be routed to the appropriate agencies and reviewers upon submission.

Project Name

Project Dates

IRB Approval

Research Question

Research Method

Expected Outcome

How will this project further the agency's mission?

Previous

Next

Navigation

Projects

Explore Data

User Directory

Bookmarks

Project Request

Please provide the following information to initiate a project request in the ADRF. Your request will be automatically be routed to the appropriate agencies and reviewers upon submission.

Overview

Members (1)

Datasets

Data Security

Select the project members from the list below

Pick new member

Graham Henke

grh255@nyu.edu

Remove

Gonen Minuskin

gm130@nyu.edu

Remove

Previous

Next

Save as Draft

Submit

Pick a dataset

Decennial Census Illinois Profile of General Population and Housing Characteristics: 2000

Dataset ID: adrf-000005

Data Steward: Drew Gordon

Remove

Previous

Next

Save as Draft

Submit

Previous

Next

Save as Draft

Submit

# Technical: Legal and Operational

ADRF | Data Stewardship

5

Michael Scott

Navigation

Projects

Explore Data

Project Review

Projects / Project Review

Overview

Members

Datasets

Agreements

Data Security

### Institutional Agreements

Agreements uploaded here are to be signed by institutional representatives only. These documents will be shared with included Data Stewards, the Principal Investigator, and Project Requester. They will **not** be shared with project members.

Attach agreement

### Individual Agreements

Agreements uploaded here will be distributed to and require signature from each member of the project. This can include Non-Disclosure Agreements, Data Use Agreements, etc.

Attach agreement

Feedback

Leave feedback

Reject

Approve

# Technical: Report

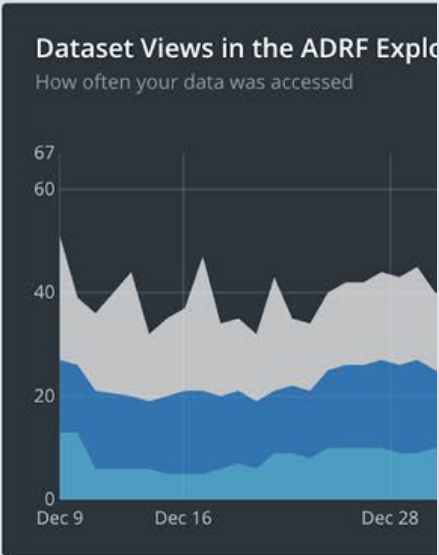
ADRF | Data Stewardship

Navigation

- Projects
- Explore Data
- User Directory
- Data Steward** NEW
- Dashboard
- Project Requests 5
- Incoming Questions
- Data Export Requests
- Bookmarks

«

## Data Steward Dashboard



ACTIVE PROJECTS

3

VIEWS

## Dataset Popularity

Dataset Name	Researchers who have access	Used in projects	Used in completed projects
National Consumer Panel - Demographics	225	38	15
National Consumer Panel - Trips	361	48	17
Med Profiler Survey	55	22	13
RX Pulse Longitudinal Panel	223	28	13
POS RMA Level	369	45	11
POS Private Label	249	24	12
POS Store Level	112	24	8
Random Weight RMA Level	319	21	13
Random Weight Store Level	195	34	5
Product Dictionary IRI	278	42	15
Random Weight Dictionary	69	38	8
Store Dictionary	313	50	24

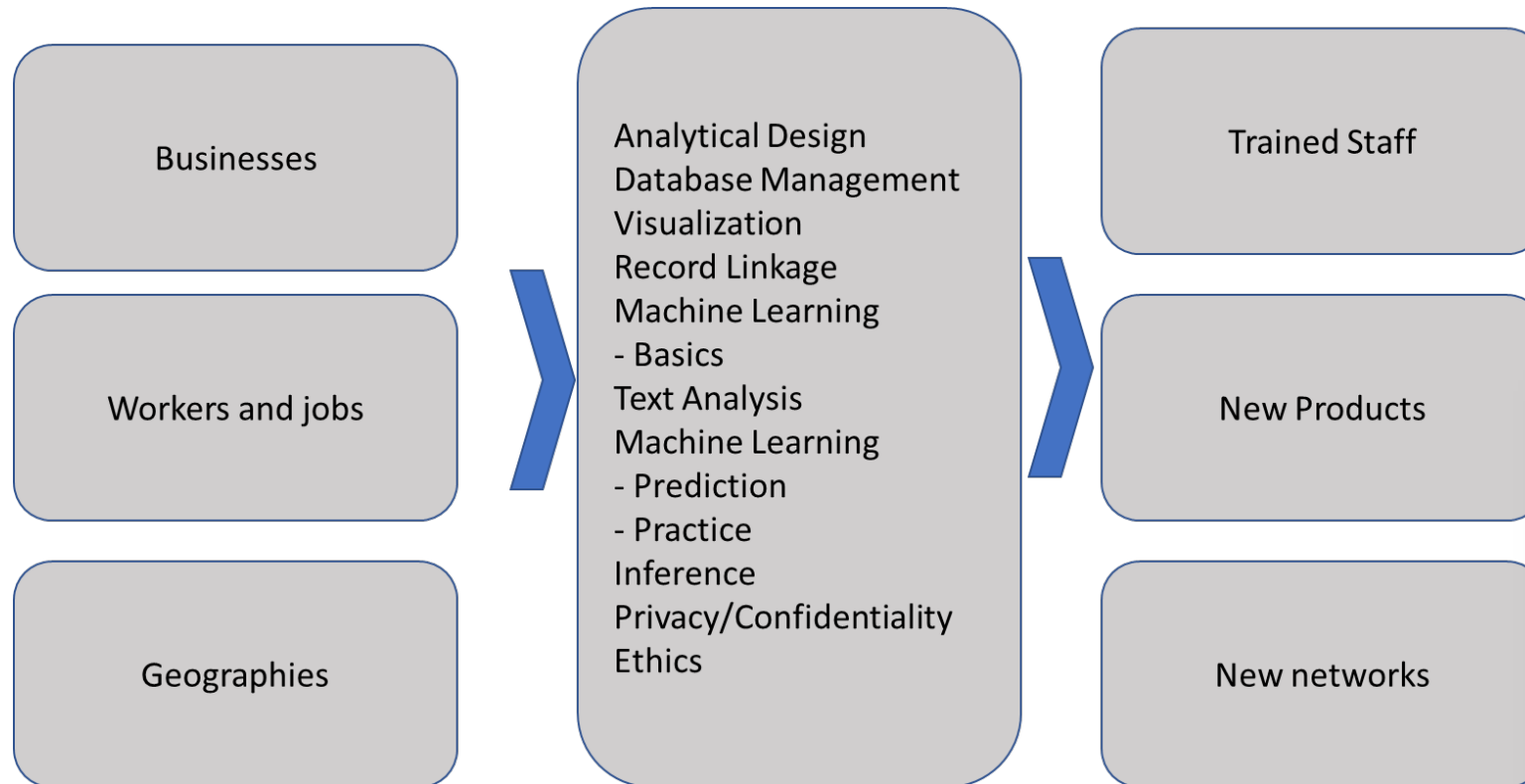
## Top Users of Your Datasets

User Name	Datasets Used
Bartholomew Kostopolous	32
Jennifer Tour Chayes	28
Grace Hopper	17
Bill Nye	12
Claudia Perlich	10

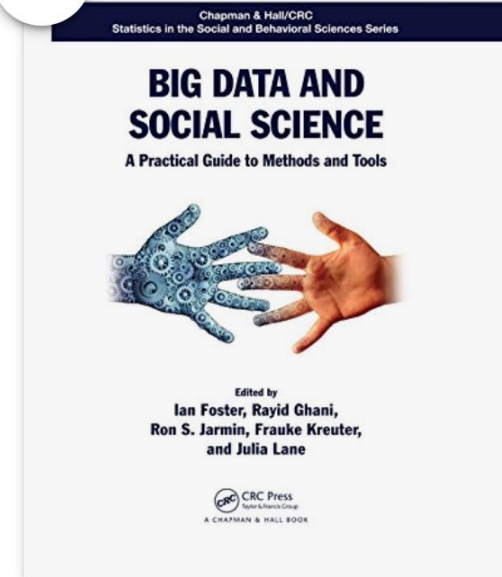
## Top Institutions Using Your Datasets

Institution	Datasets Used
Center for Urban Science + Progress	48
Chapin Hall at the University of Chicago	42
Urban Center for Computation & Data	39
Census Bureau	27
City of New York	24

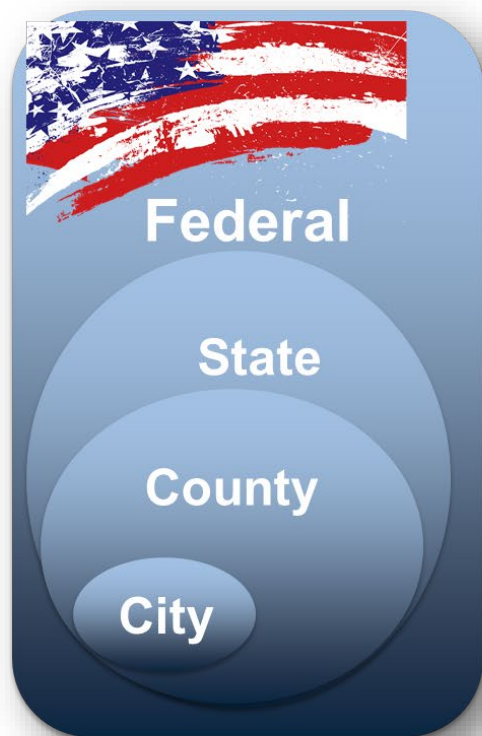
# Human: Build workforce capacity



13



# Results: Trained Staff



Over 450 participants from over 100 agencies have completed almost 100 projects

# Questions Identified (examples)

- What factors predict on-time graduation from post-secondary education?
- What factors predict post-graduation employment outcomes?
- What are the earnings and employment outcomes of various credentials and degrees, remedial skills training, vocational training, or apprenticeship training?
- To what extent are postsecondary students enrolling in majors that match to in-demand occupations?
- What proportion of postsecondary graduates work in a job that is in their field of study?
- What are the various career pathways followed by individuals?

# Class Data

Data Provider	Datasets	Example Use for Education & Work	Geographic Granularity	Time horizon & update frequency
Ohio Dept. of Job and Family Services	Wage Records	Employment, Labor Mobility (includes employer code & industry)	None	2007-2017; quarterly
Ohio Dept. of Higher Education	College student records	Enrollment and degree/certification in Ohio public, post-secondary education	County of residence	2007-2017; annually
Illinois Department of Employment Security	Employer Data Wage Records	Business Creation, Wages, Industry Trends Employment, Labor Mobility	Location of employer	2005-2015; quarterly
Missouri Department of Labor	Employer Data Wage Records	Business Creation, Wages, Industry Trends Employment, Labor Mobility	Location of employer	2006-2016; quarterly

*Complete dataset documentation here: <https://ada.coleridgeinitiative.org/osu-2019-program#data-doc>*



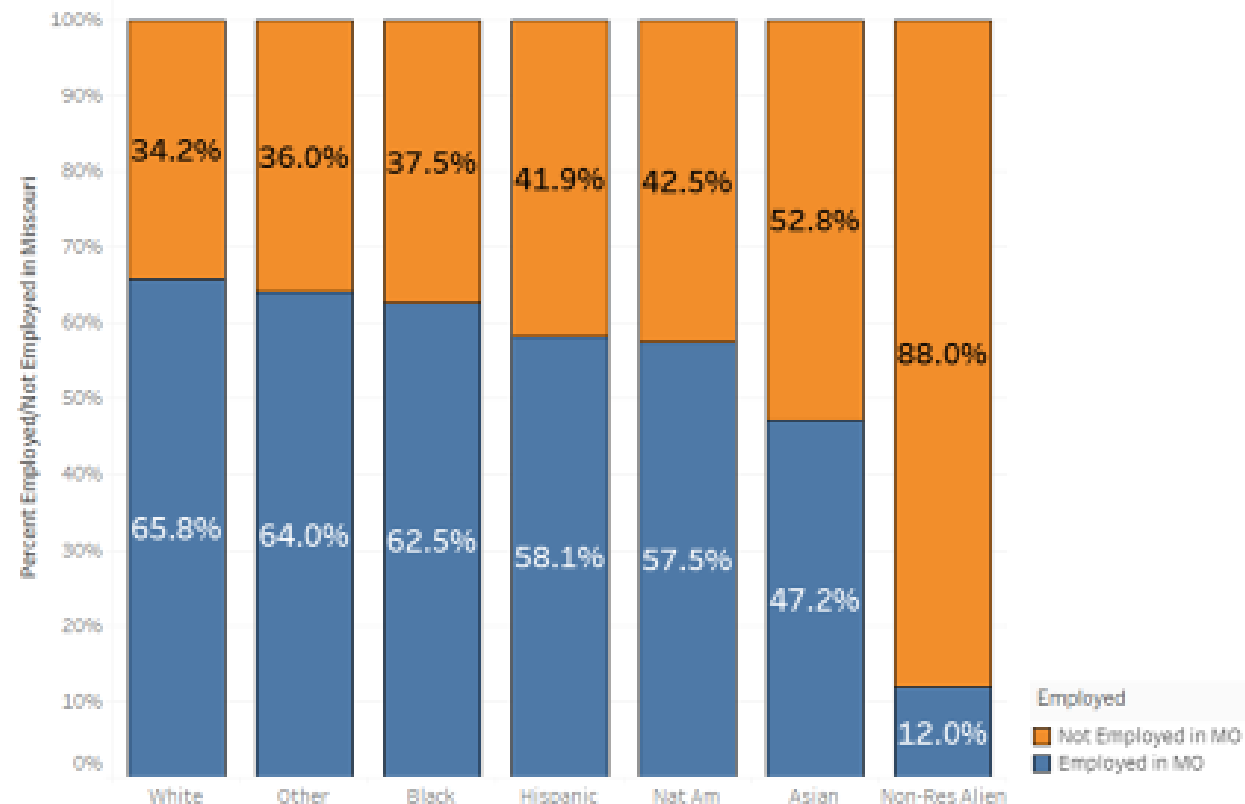
# Data continued

Data Provider	Datasets	Example Use for Education & Work	Geographic Granularity	Time horizon & update frequency
Ohio Dept. of Job and Family Services	RAPIDS	Ohio residents who participated in registered apprenticeship programs and/or participants of Ohio-based registered apprenticeship programs	N/A	2007-2017; annually
Ohio Dept. of Higher Education	Vocational training records (OTC)	Ohio Technical Center students	County	2007-2017; annually
Ohio Dept. of Higher Education	ABLE/Aspire	Ohio ABLE/Aspire students	County	2007-2017; annually
Opportunities for Ohioans with Disabilities	Vocational rehabilitation	Ohioans with disabilities employment services records	County	2011-2017; annually
Ohio Housing Finance Agency	Housing service records		Zipcode	2015-2017; annually

*Complete dataset documentation here: <https://ada.coleridgeinitiative.org/osu-2019-program#data-doc>*

# Results: Cross Agency Data

Percent Employed/Not Employed in Missouri  
1 Year After Graduation by Race  
2011 & 2012 Graduates



# Coding

- Created tables in SQL using available data.
- Queried tables to determine number of graduates and number employed/not employed.
- Categorized types of graduates to develop predictive features.
- Developed training Cohort A (2011-2012) and testing Cohort B (2013-2014) for building a machine-learning model.



# Results: Interstate data

## 1. Comparing Employment Dynamics Across Borders

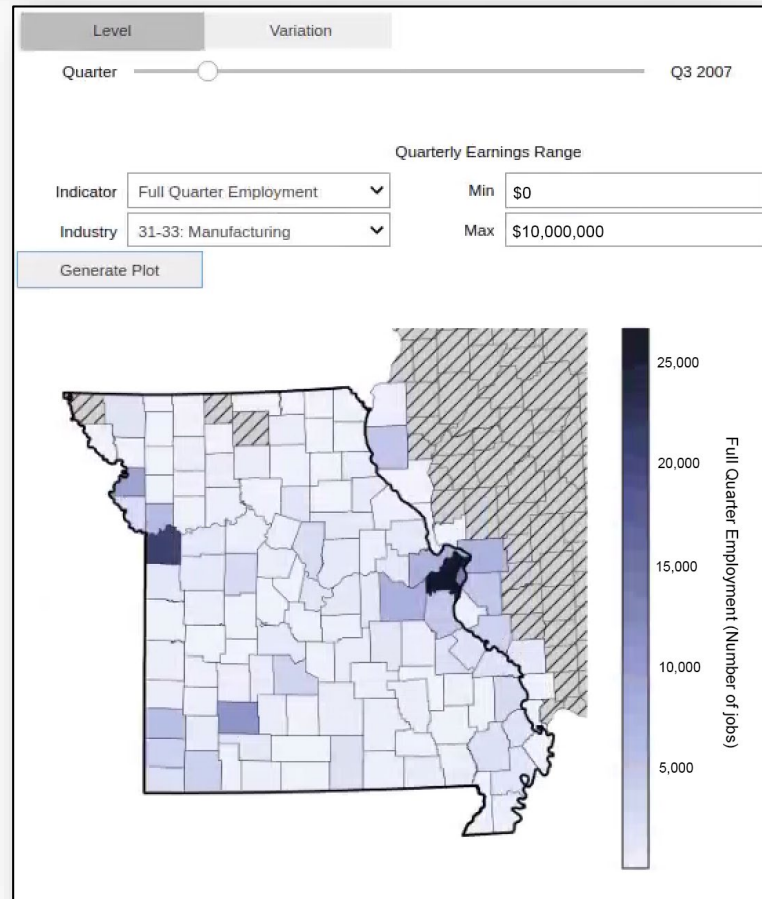


Fig. 3: Comparing total earnings with Illinois border counties

The dashboard can include border counties from the states that provide data to the ADRF.

### UPCOMING TRAINING PROGRAMS

Columbus, OH - Winter 2019

APPLY

The Ohio State University is hosting the Winter, 2019 Coleridge Initiative Applied Data Analytics training program, sponsored by the Overdeck Family Foundation. Participants will work in teams to define and complete a project related to Transitions in Education and Workforce. The program will provide up-to-date perspective on the use of administrative data for policy analysis, and hands-on instruction of using micro data in SQL and Python for the following tasks: data management, record linkage, data visualization, and machine learning.

Application: **Rolling**  
1st Admission Decision: **December 15**  
SQL & Python tutorials: Complete by **February 1**  
In-person training in Columbus, OH  
• Module 1: **February 6-8**  
• Module 2: **March 13-15**  
Remote Presentations: **April 23**

### Outcomes from Transitions in Education and Work: Developing a Scalable Regional Approach

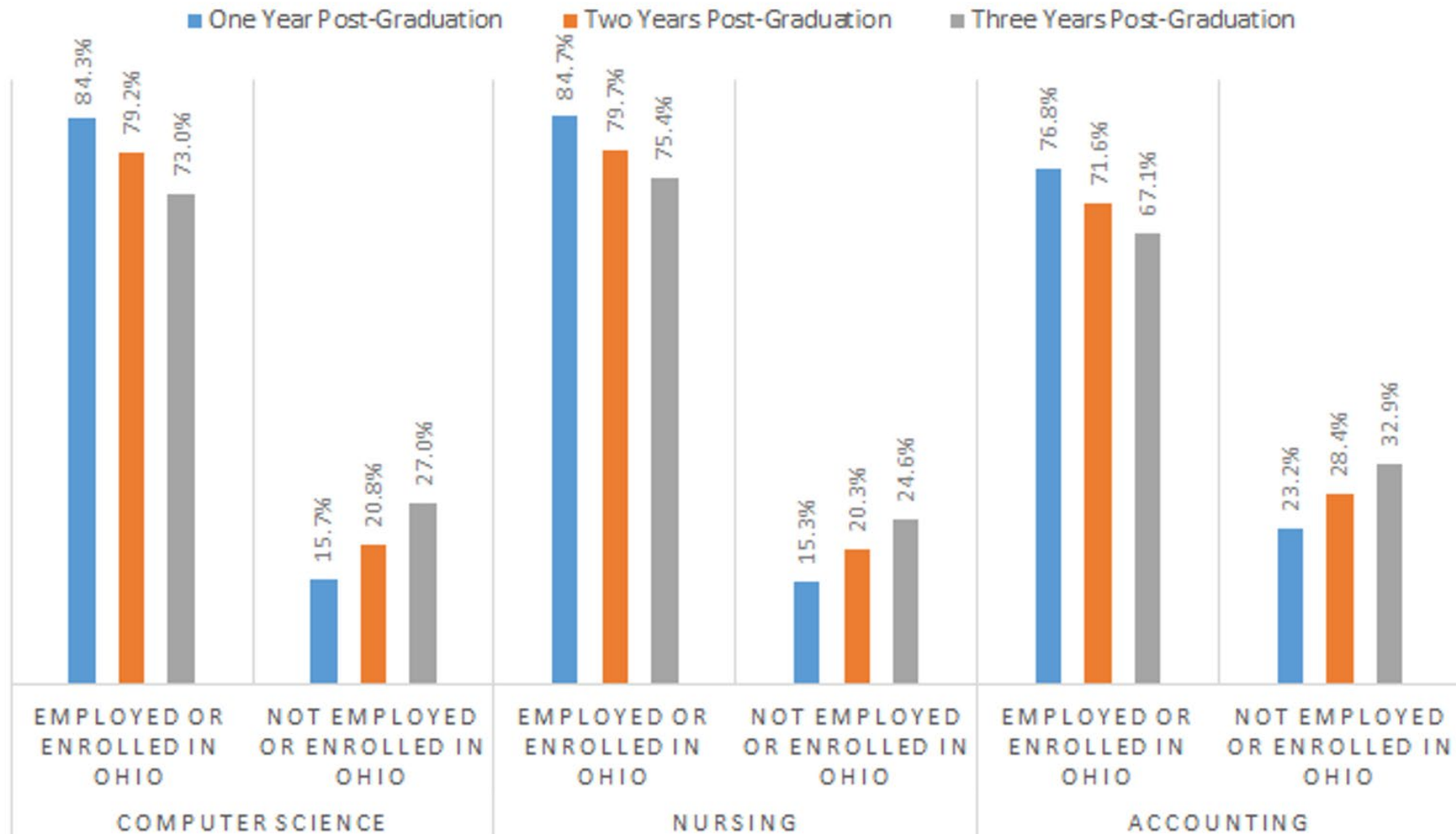
#### Summary

There is enormous interest in building a better understanding of how people transition across different educational and work experiences to sustainable jobs. Transitions of people as they age can be nonlinear and include transitions within and across secondary and post secondary institutions, as well as the use of government services such as disability services, criminal justice interventions, or workforce development. Individuals can also move across political jurisdictions such as state or county lines, making it incredibly difficult to understand the regular patterns in service utilization. The sheer variation on types of transitions individuals can make and the geographical movement across borders opens up opportunities for governments to intervene in more productive ways. The data sharing across states and agencies also provides increasing possibilities for more efficient and effective utilization of government services among vulnerable populations.

Building a common framework across states requires (i) identifying key issues of interest (ii) bringing together data from many sources across both state and agency lines and (iii) bringing together staff from those agencies. These three activities are necessary to work through the important data and measurement issues that are necessary to build operational metrics for decision-making. This workshop will build on a successful collaborative approach developed by the State of Illinois, New York University's Coleridge Initiative, Chapin Hall and Ohio State University that uses an inter-agency and inter-state collaborative, classroom environment to address the issues. It will bring together key stakeholders from the Midwest to develop an operational agenda for the spring of 2019. We expect to discuss how to develop such metrics as:

1. What are the main educational and work flows of interest for policy?
2. What data are available?
3. What kind of metrics would be of greatest use to stakeholders?

## EMPLOYMENT AND ENROLLMENT IN OHIO POST-GRADUATION



## WHAT NEXT?

- What interventions can Indiana create knowing that bachelor's graduates with the following characteristics have a higher likelihood of being employed out of state the year after graduation:
  - Earning credentials in engineering, business, computer science, comm/journalism, transportation/materials moving
  - Non-residents
  - Originally from a border state
  - Have graduation dates in quarter 2 (April-June)
  - Asian

# Example: New cross-state collaborations

## **Midwest Research Collaborative**

### Why a Midwest Collaborative?

Our common Midwestern values lead us naturally toward working together. We choose collaboration over going it alone. And what is more natural to the Midwest than data sharing and mutual research.

The Midwest has a vast network of regional activity throughout its agricultural and manufacturing base, as well as its supply chain, infrastructure and in its agile workforce, combined with hundreds of universities and colleges.

### A Shared Experience

Linked by industry and innovation, we also share regional challenges: the dynamic, changing evolution of jobs due to AI-driven development within a manufacturing-heavy economy, the effects of trade agreements, an aging population, worker shortages, migration out of rural communities, and the effects of the opioid crisis.

<https://ibrc.kelley.iu.edu/midwestcollab.html>

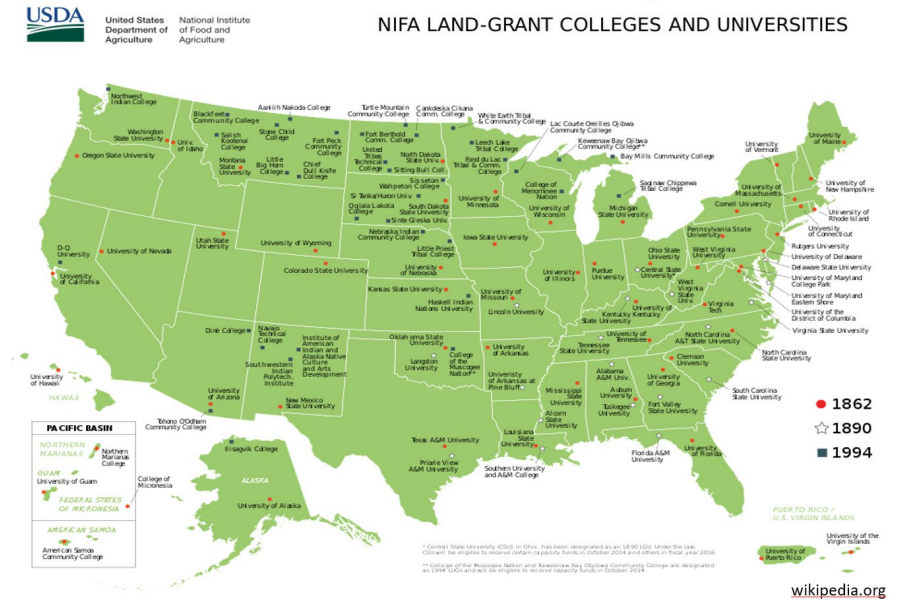


# Broader Vision

The National Academies of  
SCIENCES • ENGINEERING • MEDICINE

## CONSENSUS STUDY REPORT

# DATA SCIENCE FOR UNDERGRADUATES OPPORTUNITIES AND OPTIONS



# Key ideas

- Leverage new opportunity to use data for evidence based policy
- Build on federal and philanthropic investments
- Apply technical and human approach to create high value products
- Partner in cross-state initiative

# Contact and more information

- [Website: https://coleridgeinitiative.org/](https://coleridgeinitiative.org/)
- [E-mail](#)
  - [dataanalytics@coleridgeinitiative.org](mailto:dataanalytics@coleridgeinitiative.org)
  - [julia.lane@nyu.edu](mailto:julia.lane@nyu.edu)
  -
- [GitHub organization: https://github.com/Coleridge-Initiative](https://github.com/Coleridge-Initiative)