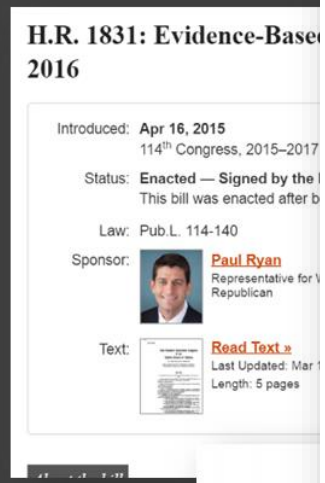
Data-Sharing and Administrative Data

JULIA LANE

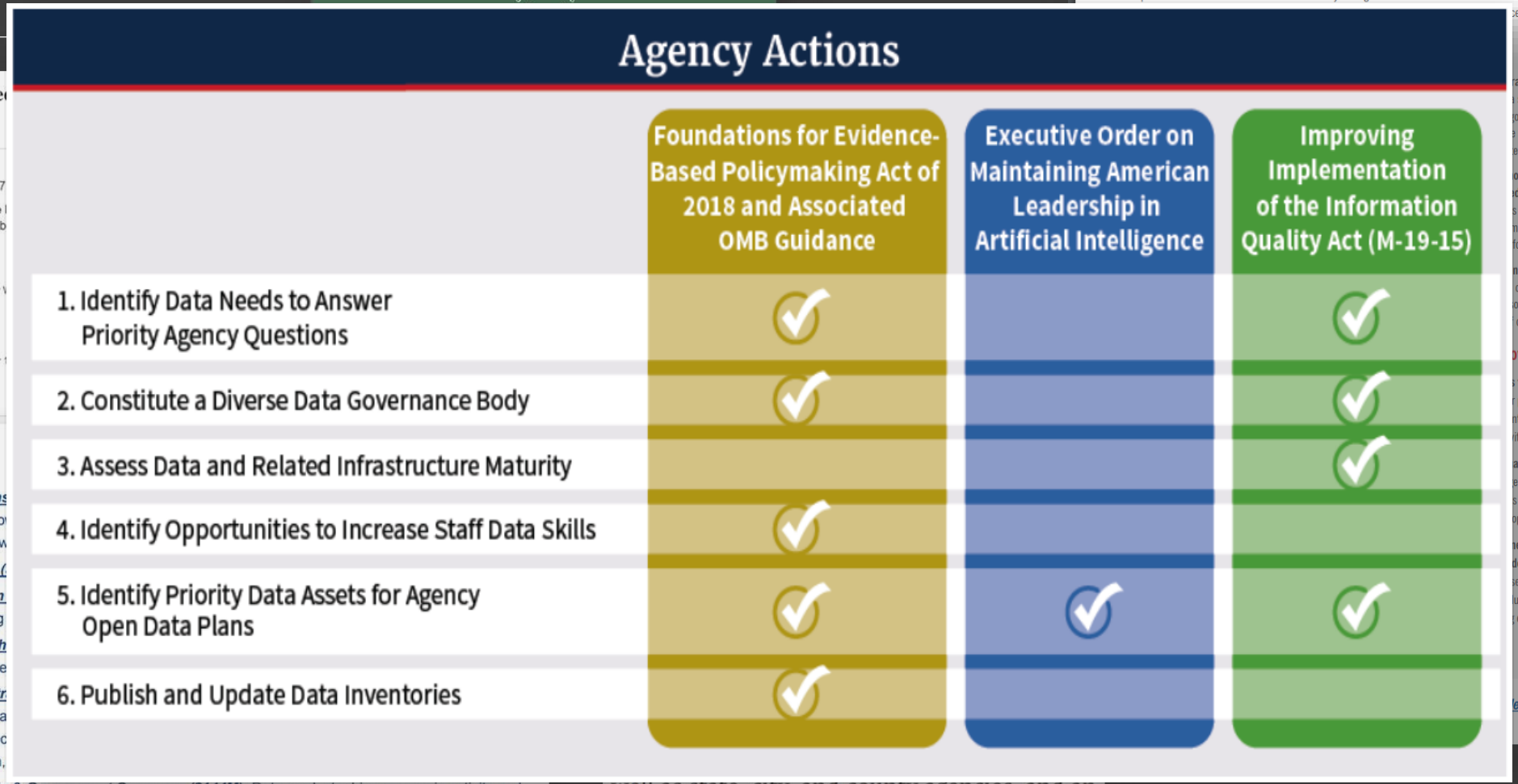
THE COLERIDGE INITIATIVE

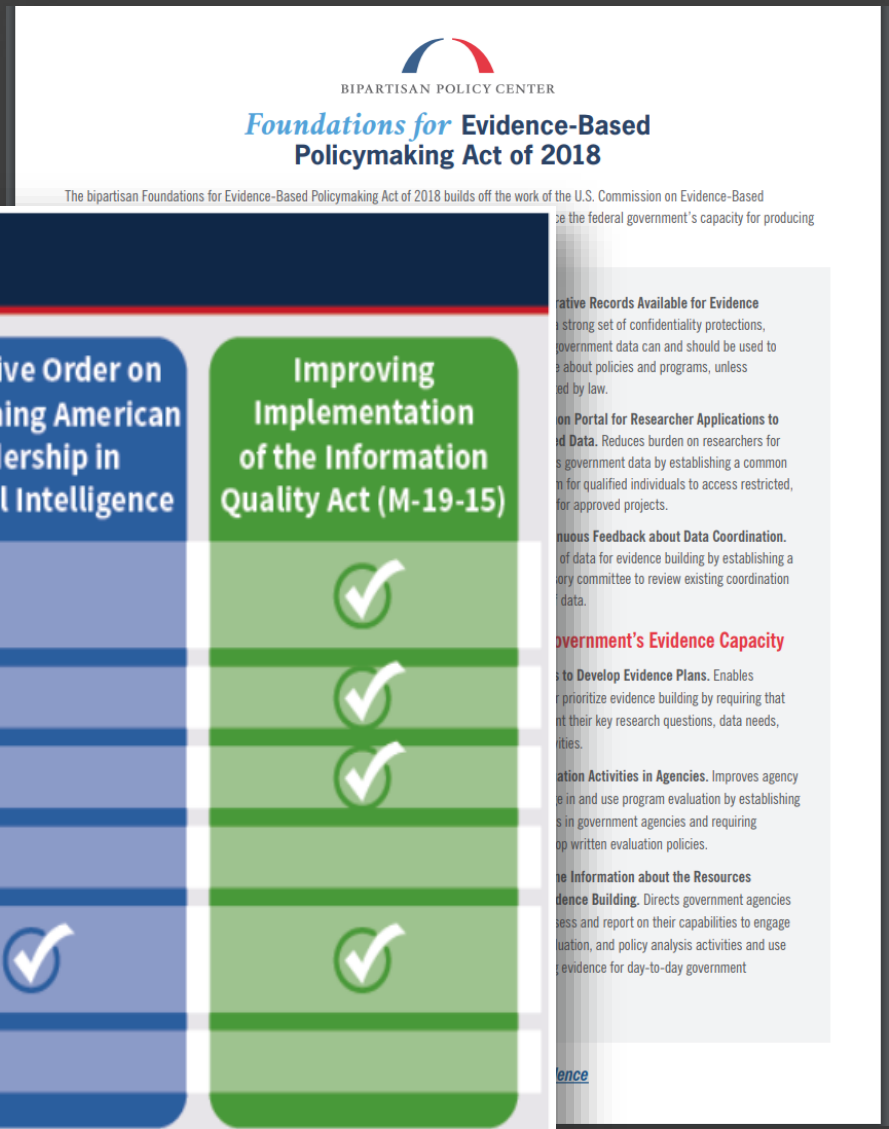
NEW YORK UNIVERSITY

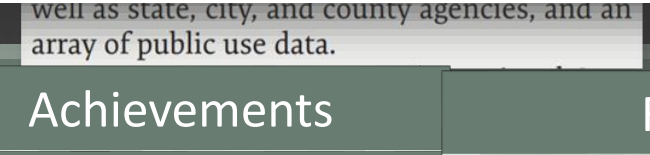
Context


H.R. 1831: Evidence-Base
2016
Introduced: Apr 16, 2015
114th Congress, 2015–2017
Status: Enacted — Signed by the
This bill was enacted after b
Law: Pub.L. 114-140
Sponsor: Paul Ryan
Representative for
Republican
Text: Read Text »
Last Updated: Mar
Length: 5 pages

2020 Cens
design, ho
on the new
CEDCaP (
American
continuing
Geograph
smarter ge
Administr
and federa
linkage tec
protection,
Economic & Government Censuses ($144M): Data products drive economic activity and
are relevant to the needs businesses, policymakers, and the public. $10.1 million increase


Agency Actions
Foundations for Evidence-Based Policymaking Act of 2018 and Associated OMB Guidance
Executive Order on Maintaining American Leadership in Artificial Intelligence
Improving Implementation of the Information Quality Act (M-19-15)
1. Identify Data Needs to Answer Priority Agency Questions
2. Constitute a Diverse Data Governance Body
3. Assess Data and Related Infrastructure Maturity
4. Identify Opportunities to Increase Staff Data Skills
5. Identify Priority Data Assets for Agency Open Data Plans
6. Publish and Update Data Inventories


BIPARTISAN POLICY CENTER
Foundations for Evidence-Based Policymaking Act of 2018
The bipartisan Foundations for Evidence-Based Policymaking Act of 2018 builds off the work of the U.S. Commission on Evidence-Based
e the federal government's capacity for producing
rative Records Available for Evidence
strong set of confidentiality protections,
government data can and should be used to
about policies and programs, unless
ed by law.
on Portal for Researcher Applications to
d Data. Reduces burden on researchers for
government data by establishing a common
for qualified individuals to access restricted,
for approved projects.
nuous Feedback about Data Coordination.
of data for evidence building by establishing a
ory committee to review existing coordination
data.
overnment's Evidence Capacity
to Develop Evidence Plans. Enables
prioritize evidence building by requiring that
nt their key research questions, data needs,
ities.
ation Activities in Agencies. Improves agency
e in and use program evaluation by establishing
s in government agencies and requiring
op written evaluation policies.
e Information about the Resources
dence Building. Directs government agencies
ess and report on their capabilities to engage
uation, and policy analysis activities and use
evidence for day-to-day government
ence


well as state, city, and county agencies, and an
array of public use data.

Approach: Technical and Human

Technical

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

Operational

Link disparate data

Analyze data

Legal & Practical

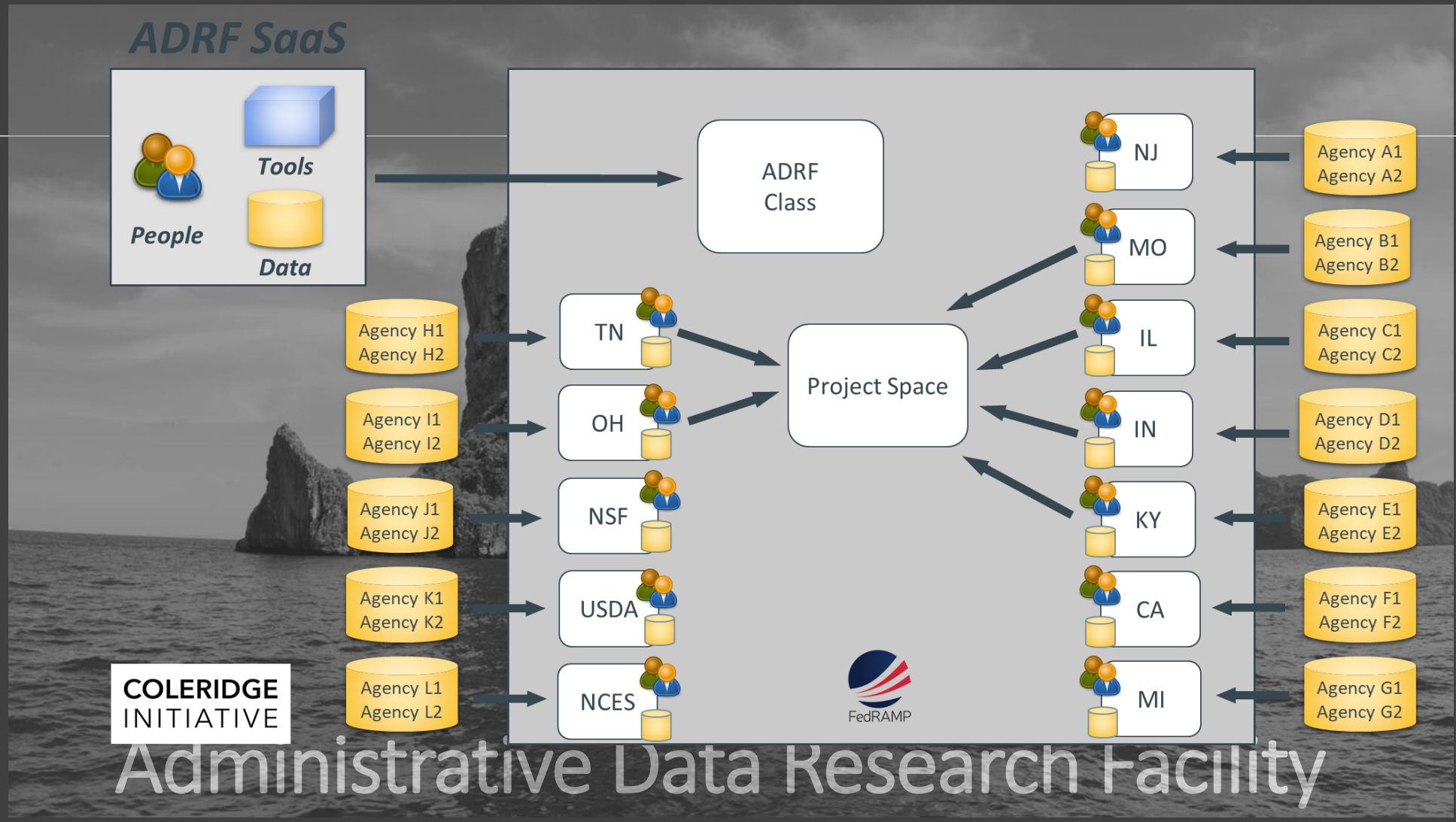
- Document value associated with the data linkage
  - Consistent with the agency mission
  - Useful enough to engage decision-makers

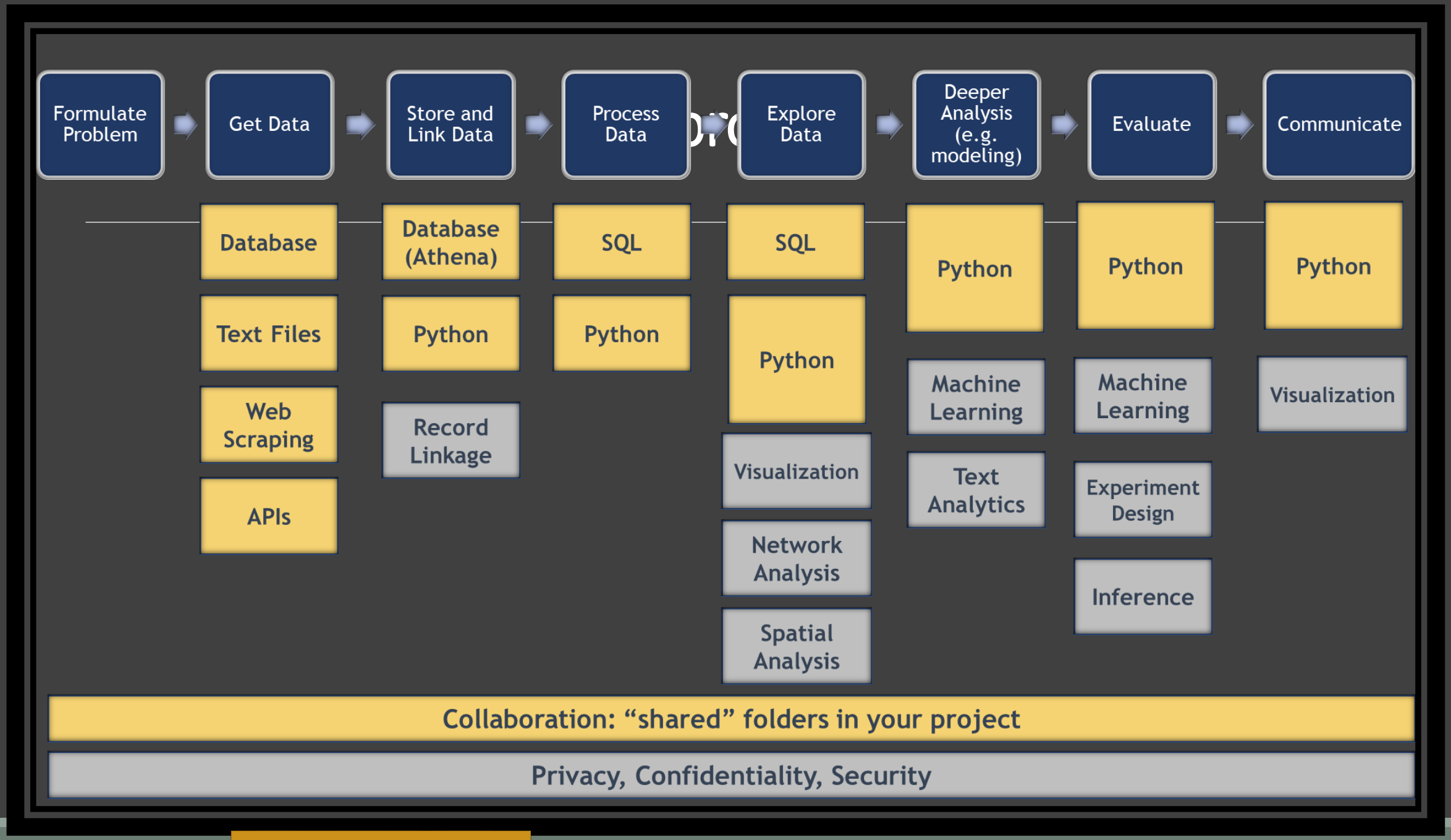
Technical Approach: Five Safes

- 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

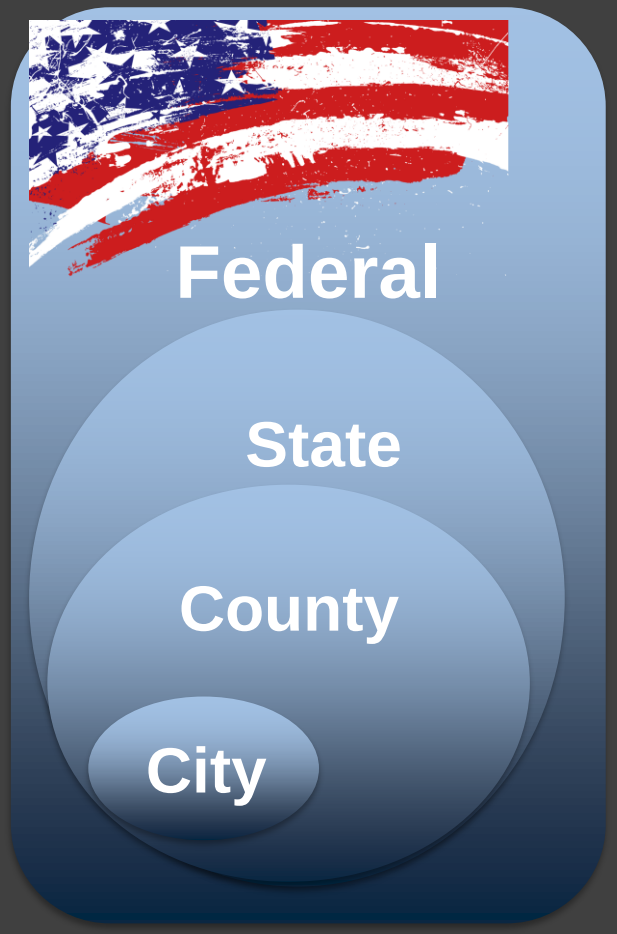
= SAFE USE

Safe Setting

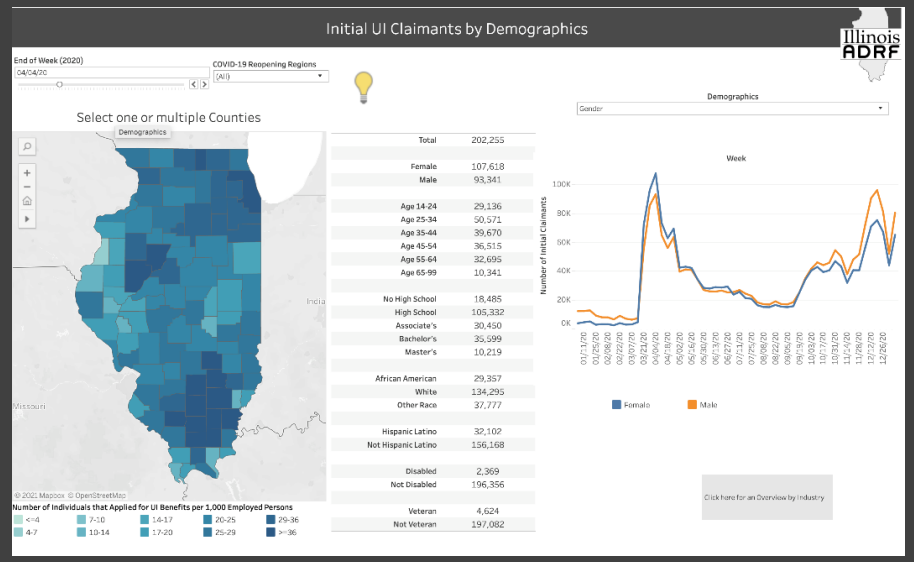


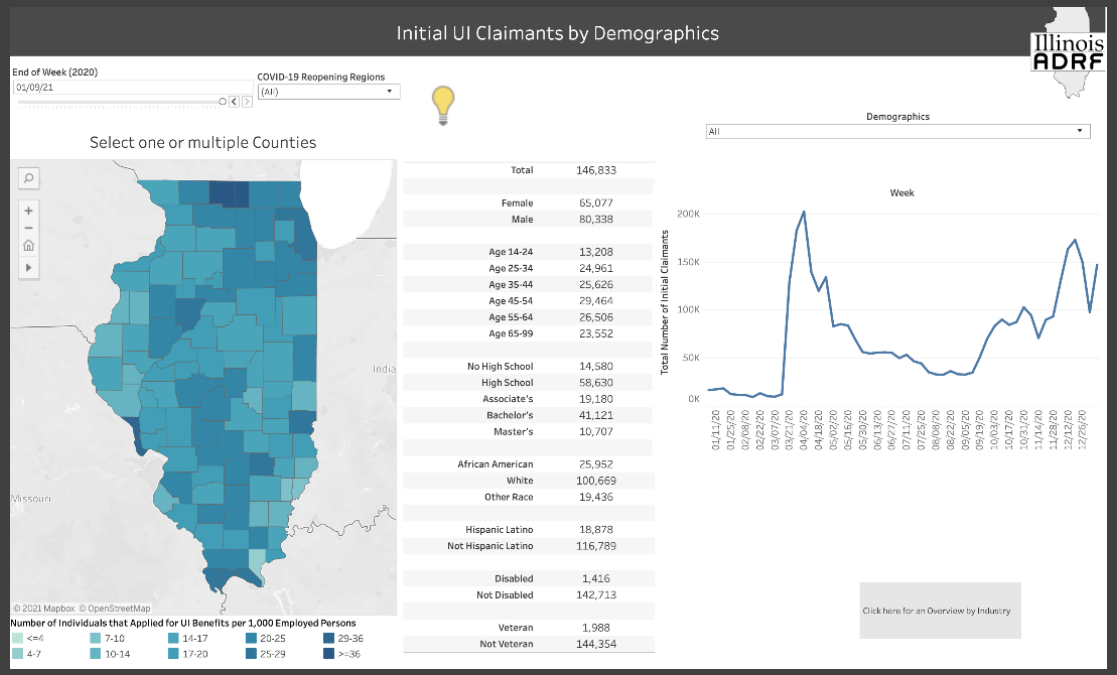

Formulate Problem
Get Data
Store and Link Data
Process Data
Explore Data
Deeper Analysis (e.g. modeling)
Evaluate
Communicate
Database
Text Files
Web Scraping
APIs
Database (Athena)
Python
Record Linkage
SQL
Python
SQL
Python
Visualization
Network Analysis
Spatial Analysis
Python
Machine Learning
Text Analytics
Python
Machine Learning
Experiment Design
Inference
Python
Visualization
Collaboration: “shared” folders in your project
Privacy, Confidentiality, Security

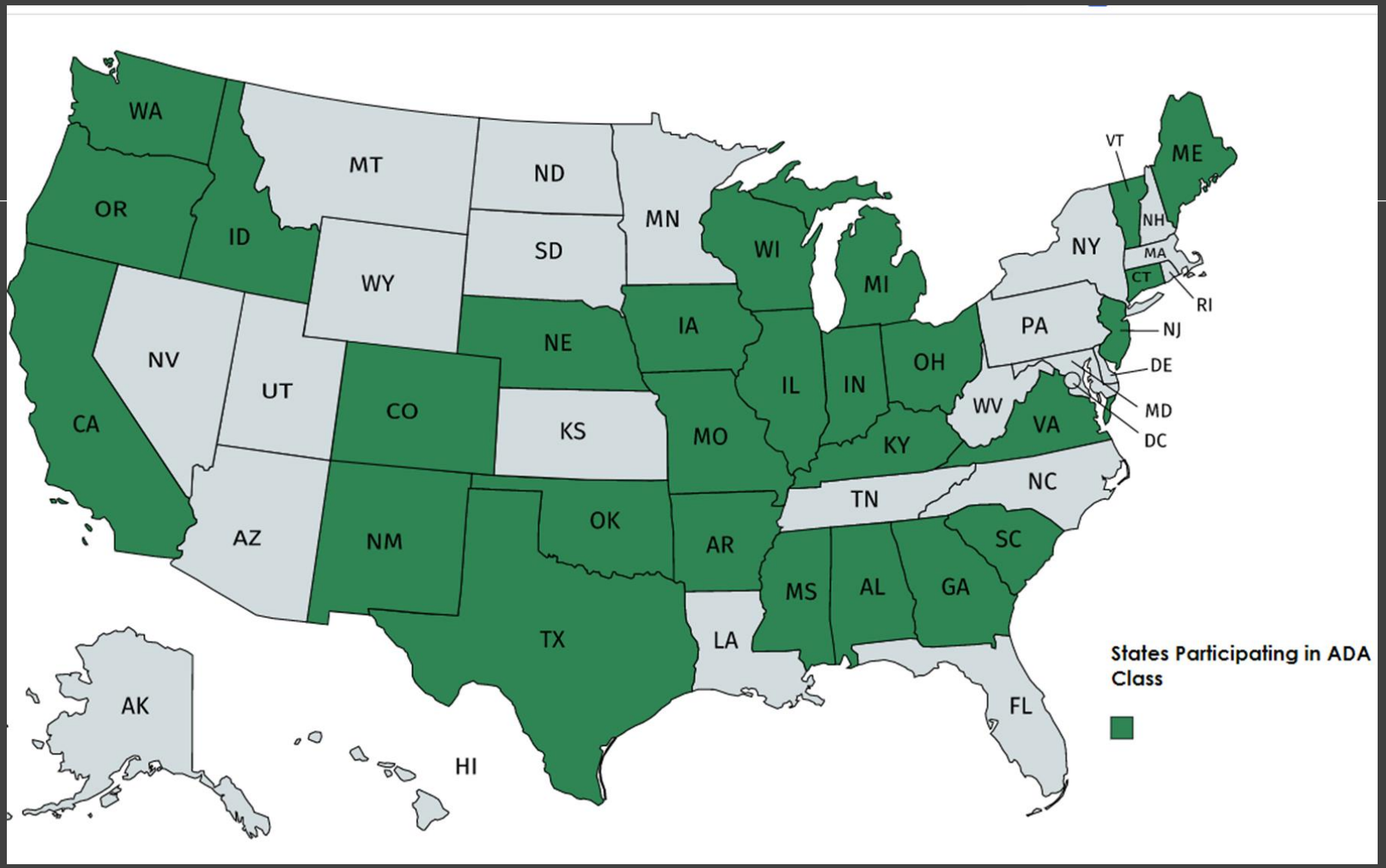
Work in teams

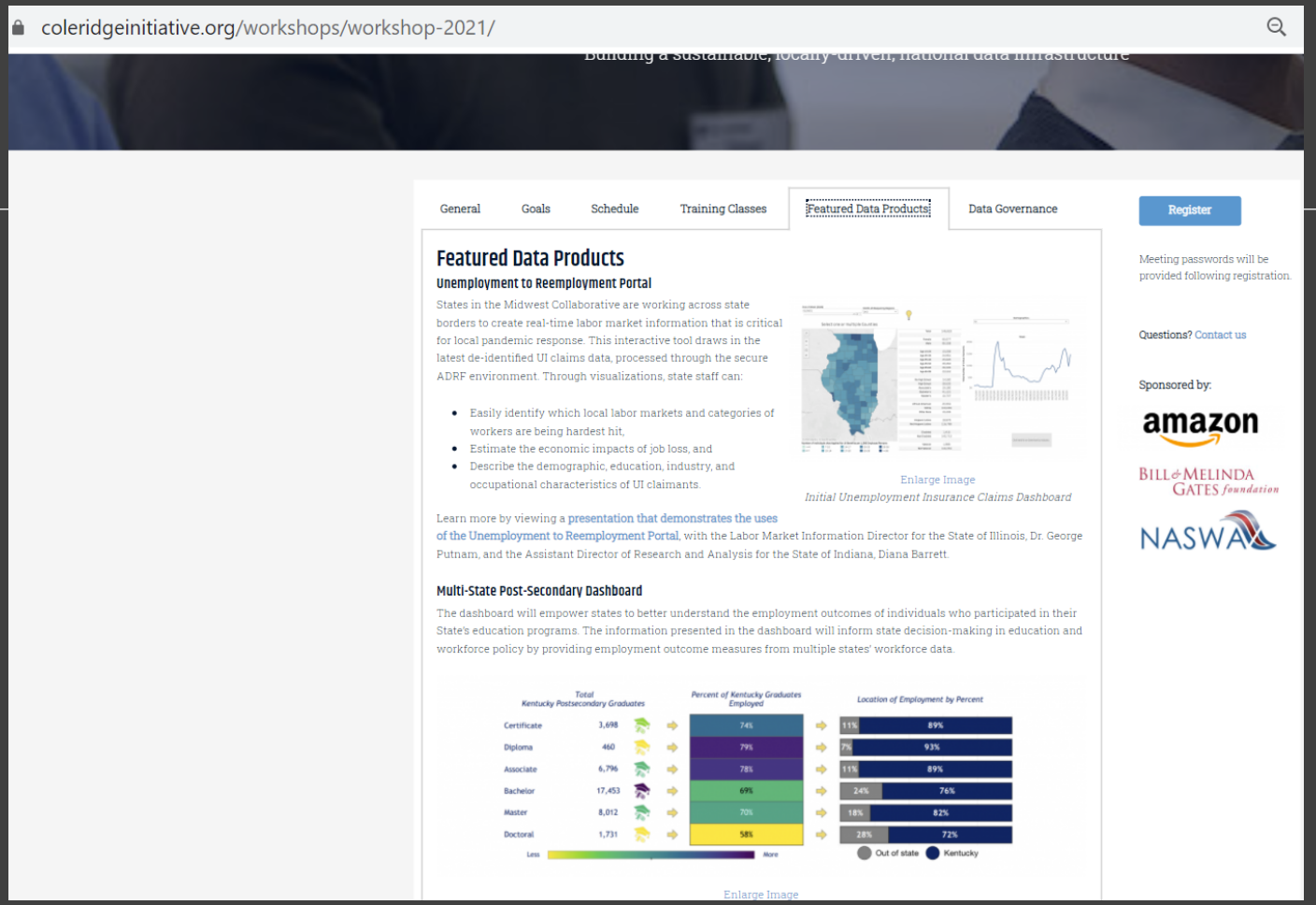


Unemployment to Reemployment Portal: Demographics






WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
VT
NH
ME
MA
CT
RI
NJ
DE
MD
DC
AK
HI
States Participating in ADA Class

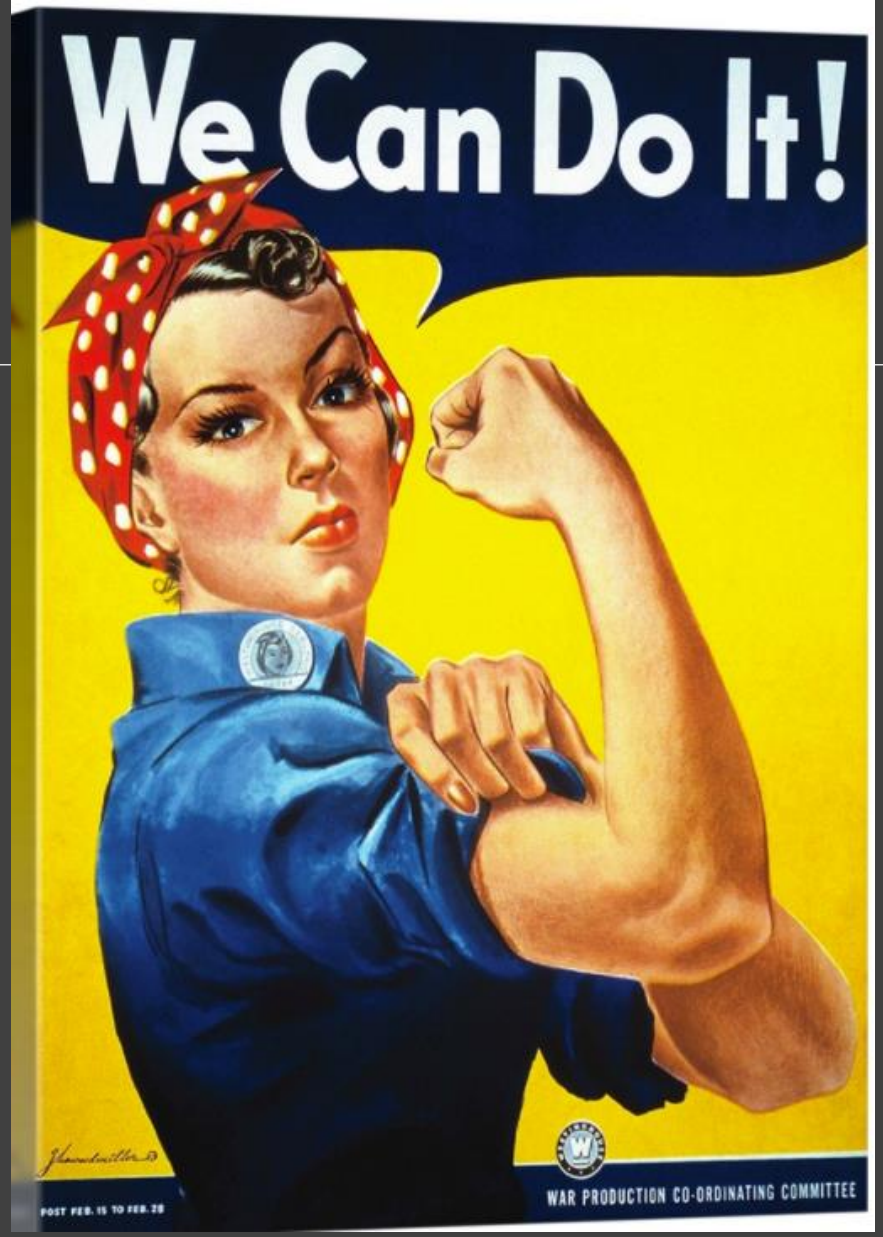

coleridgeinitiative.org/workshops/workshop-2021/
General
Goals
Schedule
Training Classes
Featured Data Products
Data Governance
Register
Featured Data Products
Unemployment to Reemployment Portal
States in the Midwest Collaborative are working across state borders to create real-time labor market information that is critical for local pandemic response. This interactive tool draws in the latest de-identified UI claims data, processed through the secure ADRF environment. Through visualizations, state staff can:
Easily identify which local labor markets and categories of workers are being hardest hit,
Estimate the economic impacts of job loss, and
Describe the demographic, education, industry, and occupational characteristics of UI claimants.
Enlarge Image
Initial Unemployment Insurance Claims Dashboard
Learn more by viewing a presentation that demonstrates the uses of the Unemployment to Reemployment Portal, with the Labor Market Information Director for the State of Illinois, Dr. George Putnam, and the Assistant Director of Research and Analysis for the State of Indiana, Diana Barrett.
Multi-State Post-Secondary Dashboard
The dashboard will empower states to better understand the employment outcomes of individuals who participated in their State's education programs. The information presented in the dashboard will inform state decision-making in education and workforce policy by providing employment outcome measures from multiple states' workforce data.
Total Kentucky Postsecondary Graduates
Percent of Kentucky Graduates Employed
Location of Employment by Percent
Certificate 3,698 74% 11% 89%
Diploma 460 79% 7% 93%
Associate 6,796 78% 11% 89%
Bachelor 17,453 69% 24% 76%
Master 8,012 70% 18% 82%
Doctoral 1,731 58% 28% 72%
Less
More
Out of state
Kentucky
Enlarge Image
Meeting passwords will be provided following registration.
Questions? Contact us
Sponsored by:
amazon
BILL & MELINDA GATES foundation
NASWA

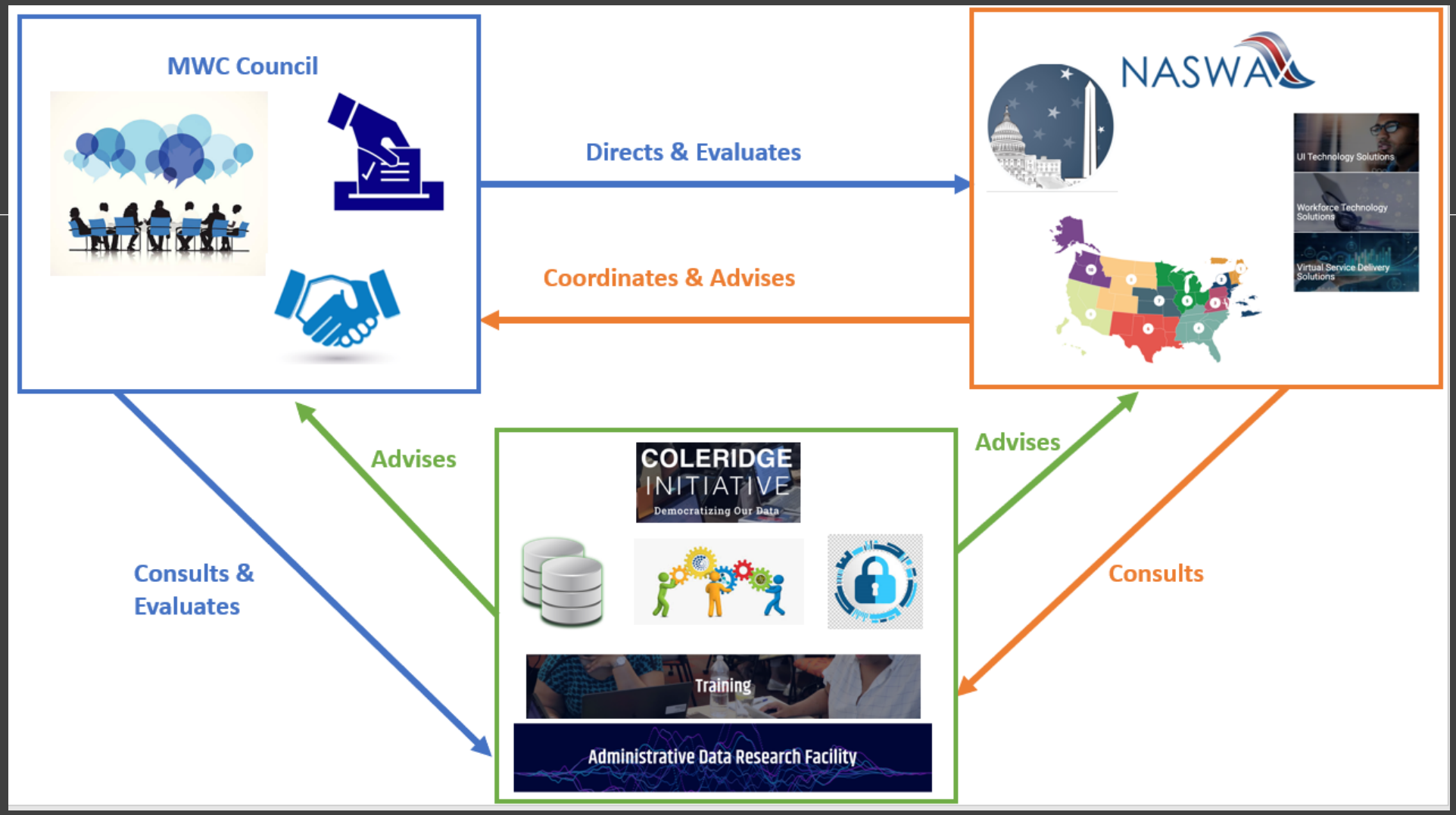
COVID-19 and the resulting
budget crisis have heightened
the importance of training an
improve outcomes for Americ
at the forefront of making trai
decisions — and they need be
evidence to make wise decisio

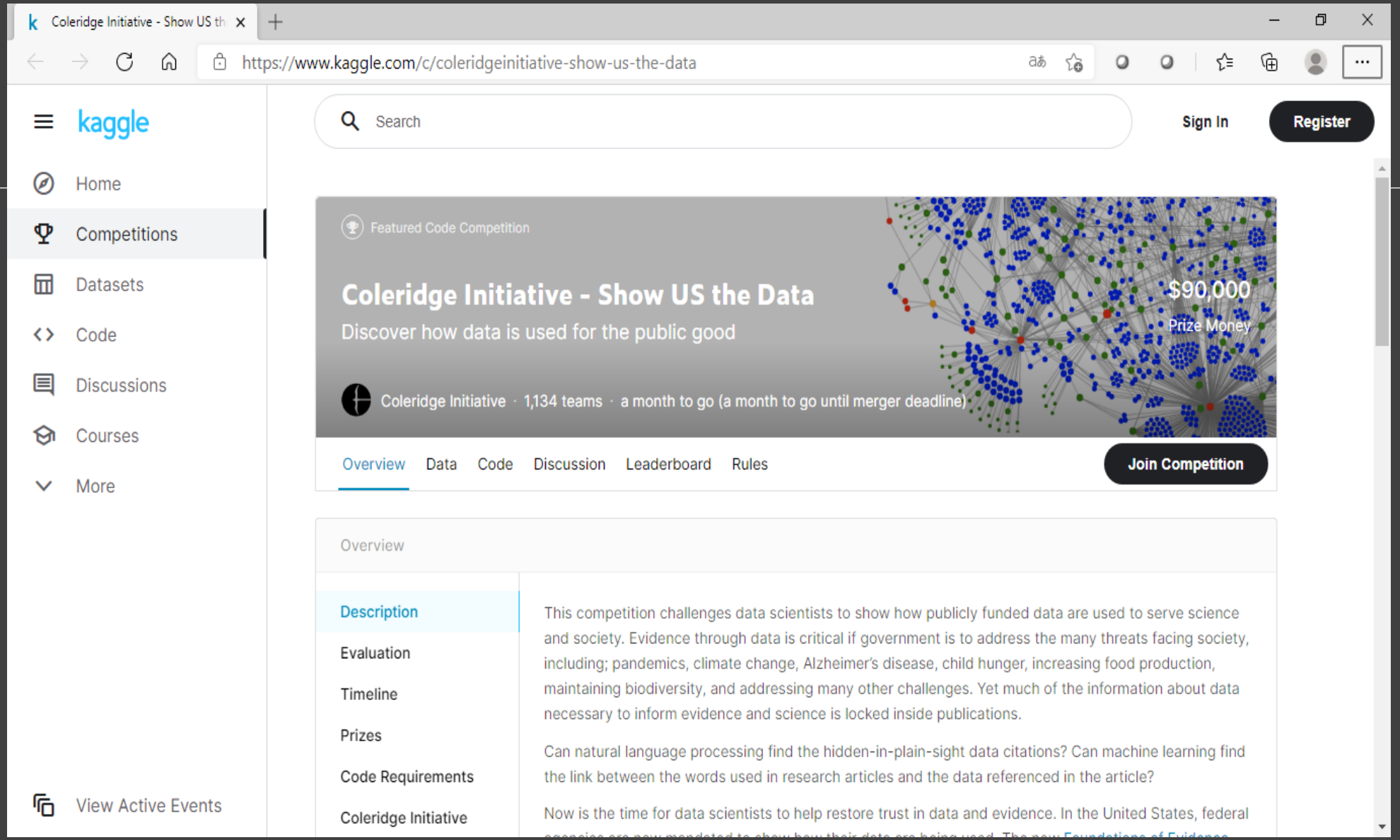
The Multi-State Post-Secondary Dashboard

Unlisted

Context | Approach | Achievements | Future | Lessons Learned


We Can Do It!
WAR PRODUCTION CO-ORDINATING COMMITTEE


MWC Council
Directs & Evaluates
Coordinates & Advises
NASWA
UI Technology Solutions
Workforce Technology Solutions
Virtual Service Delivery Solutions
Advises
Advises
Consults & Evaluates
Consults
COLERIDGE INITIATIVE
Democratizing Our Data
Training
Administrative Data Research Facility


Coleridge Initiative - Show US th
https://www.kaggle.com/c/coleridgeinitiative-show-us-the-data
kaggle
Search
Sign In
Register
Home
Competitions
Datasets
Code
Discussions
Courses
More
View Active Events
Featured Code Competition
Coleridge Initiative - Show US the Data
Discover how data is used for the public good
$90,000
Prize Money
Coleridge Initiative · 1,134 teams · a month to go (a month to go until merger deadline)
Overview
Data
Code
Discussion
Leaderboard
Rules
Join Competition
Overview
Description
Evaluation
Timeline
Prizes
Code Requirements
Coleridge Initiative
This competition challenges data scientists to show how publicly funded data are used to serve science and society. Evidence through data is critical if government is to address the many threats facing society, including; pandemics, climate change, Alzheimer's disease, child hunger, increasing food production, maintaining biodiversity, and addressing many other challenges. Yet much of the information about data necessary to inform evidence and science is locked inside publications.
Can natural language processing find the hidden-in-plain-sight data citations? Can machine learning find the link between the words used in research articles and the data referenced in the article?
Now is the time for data scientists to help restore trust in data and evidence. In the United States, federal

Quick wins and visibility key (training programs have been magic)

Security necessary but not sufficient

Strong, trusted leadership

Value proposition is key

- Create value for agencies -
- Operational and implementable (timeliness often key)
- Analysis of outliers
- Analysis of small populations
- Analysis of regions and localities

Workforce capacity is critical

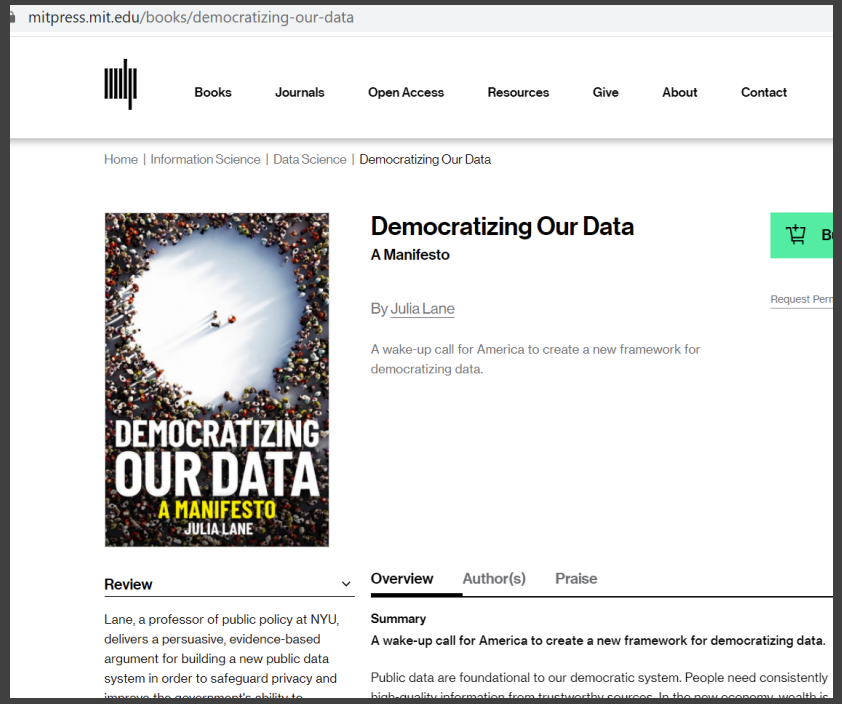
Long term investment necessary

https://hdsr.mitpress.mit.edu/pub/0mb0zzlc/release/5

https://www.nature.com/articles/d41586-020-02733-3

https://www.scientificamerican.com/article/the-u-s-needs-a-national-data-service/

https://gcn.com/Articles/2018/11/01/PSI_Administrative-Data-Research-Facility.aspx?m=1



Contact and more information

Website: https://coleridgeinitiative.org/

Email
- julia.lane@coleridgeinitiative.org

GitHub organization: https://github.com/Coleridge-Initiative

Class information: https://ada.coleridgeinitiative.org (password adrf)

Textbook: https://textbook.coleridgeinitiative.org/

Kaggle competition: https://www.kaggle.com/c/coleridgeinitiative-show-us-the-data