



Democratizing our Data





Overview

- Context
- Secure Environment
- Workforce Capacity
- Search and Discovery
- Questions



H.R. 1831: Evidence-Based Policymaking Commission Act of 2016

Introduced: **Apr 16, 2015**
114th Congress, 2015–2017

Status: **Enacted — Signed by the President on Mar 30, 2016**
This bill was enacted after being signed by the President on March 30, 2016.

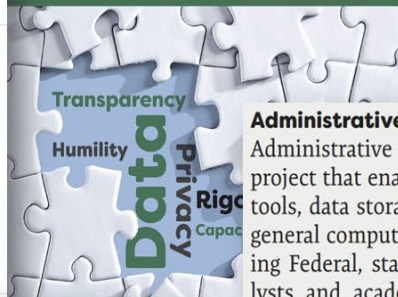
Law: Pub.L. 114-140

Sponsor:  **Paul Ryan**
Representative for Wisconsin's 1st congressional district
Republican

Text:  **Read Text »**
Last Updated: Mar 18, 2016
Length: 9 pages

THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking



FY 2016 Significant Investments

- **2020 Census (\$663M):** We have the potential to save \$5 billion with the new 2020 Census design, however, we now have to build operations and systems for the 2020 Census, based on the new design.
- **CEDCaP (\$78M):** Smarter-IT Delivery Built on a Shared-Services Model.
- **American Community Survey (\$257M):** We must maintain the quality of the data while continuing our efforts to reduce respondent burden.
- **Geographic Support (\$81M):** We must make use of technology and partnerships to deliver smarter geographic solutions to our surveys and censuses.
- **Administrative Records Clearinghouse (\$10M):** Will expedite the acquisition of federal and federally sponsored administrative data sources, improve data documentation and linkage techniques, and leverage and extend existing systems for governance, privacy protection, and secure access to these data.
- **Economic & Government Censuses (\$144M):** Data products drive economic activity and are relevant to the needs businesses, policymakers, and the public. \$10.1 million increase

Administrative Data Research Facility: The Administrative Data Research Facility is a pilot project that enables secure access to analytical tools, data storage and discovery services, and general computing resources for users, including Federal, state, and local government analysts and academic researchers. The Census Bureau and academic partners developed the project as part of the collaborative Training Program in Applied Data Analytics sponsored by the University of Chicago, New York University, and the University of Maryland.¹ It is currently operating as a pilot with users accessing the Facility as part of the training program. The Facility operates as a cloud-based computing environment, with Federal security approvals, which currently hosts selected confidential data from the U.S. Department of Housing and Urban Development and the Census Bureau, as well as state, city, and county agencies, and an array of public use data.



Coleridge Initiative



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 Policymaking to strengthen data privacy protections, improve secure access to data, and enhance the federal government's capacity for producing and using evidence.

Strengthens Privacy Protections

Maintains Strong Confidentiality Protections for Sensitive Data. Reauthorizes the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), an existing law that gives the American public strong privacy safeguards and legal protections for appropriate uses of confidential data.

Institutes Processes to Assess Data Risks. Strengthens efforts to protect confidentiality while making data accessible for evidence building and transparent to the public by requiring comprehensive risk assessments for certain publicly released data.

Enhances Public Trust in Data. Improves public trust in statistical activities by codifying language directing certain agencies to establish procedures to protect trust in data activities by appropriately maintaining objectivity, independence, and confidentiality.

Establishes Consistent Leadership on Key Data Issues. Ensures a senior leader in each agency is responsible for protecting privacy and ensuring confidentiality protections are appropriately applied by creating chief data officers.

Improves Secure Data Access

Encourages Agencies to Make Data Public and Open When Possible. Takes steps to improve the public information about what data government currently holds and make data publicly available when possible and in the public interest.

Requires Development of Data Inventories. Enables researchers and evaluators to better identify what government-collected data are available by directing agencies to create and maintain data inventories and publicly provide details about those datasets.

Makes Administrative Records Available for Evidence Building. Under a strong set of confidentiality protections, encourages that government data can and should be used to generate evidence about policies and programs, unless otherwise restricted by law.

Creates a Common Portal for Researcher Applications to Access Restricted Data. Reduces burden on researchers for applying to access government data by establishing a common application system for qualified individuals to access restricted, confidential data for approved projects.

Facilitates Continuous Feedback about Data Coordination. Promotes the use of data for evidence building by establishing a government advisory committee to review existing coordination and availability of data.

Enhances Government's Evidence Capacity

Directs Agencies to Develop Evidence Plans. Enables agencies to better prioritize evidence building by requiring that agencies document their key research questions, data needs, and planned activities.

Prioritizes Evaluation Activities in Agencies. Improves agency capacity to engage in and use program evaluation by establishing evaluation officers in government agencies and requiring agencies to develop written evaluation policies.

Develops Baseline Information about the Resources Available for Evidence Building. Directs government agencies to periodically assess and report on their capabilities to engage in statistical, evaluation, and policy analysis activities and use the corresponding evidence for day-to-day government operations.

Learn more at bipartisanpolicy.org/evidence

M-21-27

MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young
Acting Director

SUBJECT: Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans

The Foundations for Evidence-Based Policymaking Act of 2018¹ (Evidence Act) urges the Federal Government to make decisions using the best available evidence. The complex issues and challenges facing the American people must be met with urgency, and doing so requires the use of facts arrived at through rigorous and systematic analysis, governed by principles of scientific integrity. In order to address these issues, it is critical to ensure, protect, and institutionalize the collection, dissemination, and use of high-quality evidence in a way that is informed by diverse viewpoints and methods. Addressing and solving current national crises, such as the COVID-19 pandemic or economic downturn, as well as future crises, depends on using the best available science and evidence. This guidance responds to the Presidential Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking.² It reaffirms and expands on previous OMB guidance on Learning Agendas and Annual Evaluation Plans, including OMB M-19-23,³ OMB M-20-12,⁴ and OMB Circular A-11.⁵

This guidance applies to all agencies; CFO Act agencies have a statutory requirement as described in Title I of the Evidence Act,⁷ and developing Learning Agendas and Annual Evaluation Plans benefits all agencies at both the agency and sub-agency levels. It is only through this shift to a culture of evidence, supported and demanded by agency leaders and brought to bear across agency functions, that we will build and maintain trust in government and ensure that decisions best serve the American people.

Open Data

- The "OPEN Government Data Act" became Title II of the Foundations for Evidence-Based Policymaking Act [PL 115-435]
 - This Act amended Section 3504(b) of title 44, United States Code to make **data open by default** which includes guidance to
 - Make each data asset of the agency available in an open format
 - Make each public data asset of the agency available as an open Government data asset under an open license
 - Ensure that any public data asset of the agency is machine-readable
 - Engage the public in using public data assets of the agency
 - Publish on the website of the agency, on a regular basis (not less than annually), information on the usage of such assets by non-Government users
 - Provide the public with the opportunity to request specific data assets to be prioritized for disclosure and to provide suggestions for the development of agency criteria with respect to prioritizing data assets for disclosure
 - Assist the public in expanding the use of public data assets by hosting challenges, competitions, events, or other initiatives





Overview

- Context
- **Secure Environment**
- Workforce Capacity
- Search and Discovery
- Questions



Coleridge Initiative

Secure Environment Approach

Agency

- Security: Five Safes
- Reporting: Data Stewardship
- Transparency: Pricing and Usage

Researcher

- Ease of use: Windows and Tools
- Customization: Choice
- Collaboration: Rich Context



Approach

Agency

- Security: Five Safes
- Reporting: Data Stewardship
- Transparency: Pricing and Usage

Researcher

- Ease of use: Windows and Tools
- Customization: Choice
- Collaboration: Rich Context



Coleridge Initiative

Security: Five Safes

Five Safes Data Strategy

The ADRF platform implements the Five Safes Framework to ensure the safe and responsible use of data. To learn more about each of the five safes, click on the appropriate button below.



Safe Projects



Safe People



Safe Settings



Safe Data



Safe Exports

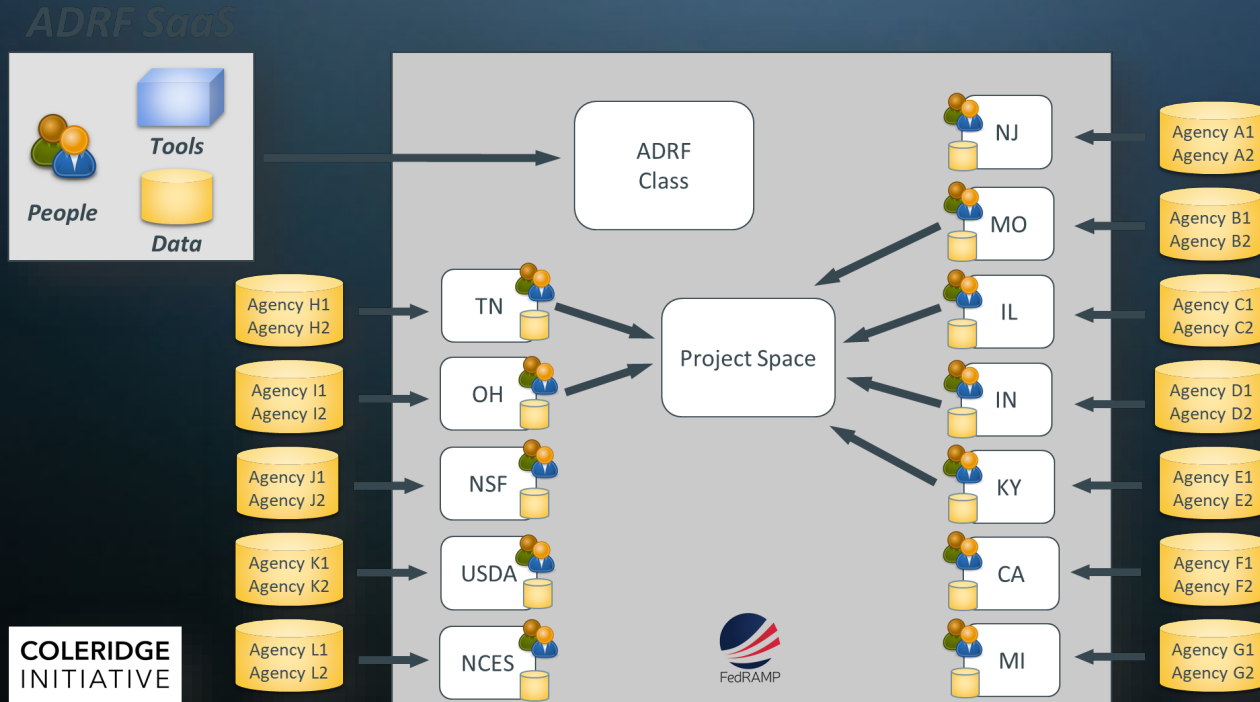


Approach

- NCSES: Survey of Earned Doctorates; Survey of Doctorate Recipients: Researchers: 20
- NCES: All 15 survey datasets Researchers: currently 50; increasing to up to 4,000
- USDA: ERS and NASS: All 21 survey and census data Researchers: increasing to up to 300
- HHS: National Directory of New Hires ; National TANF dataset
- Non Federal Data
- IRIS/UMETRICS All data: Researchers increasing up to 200



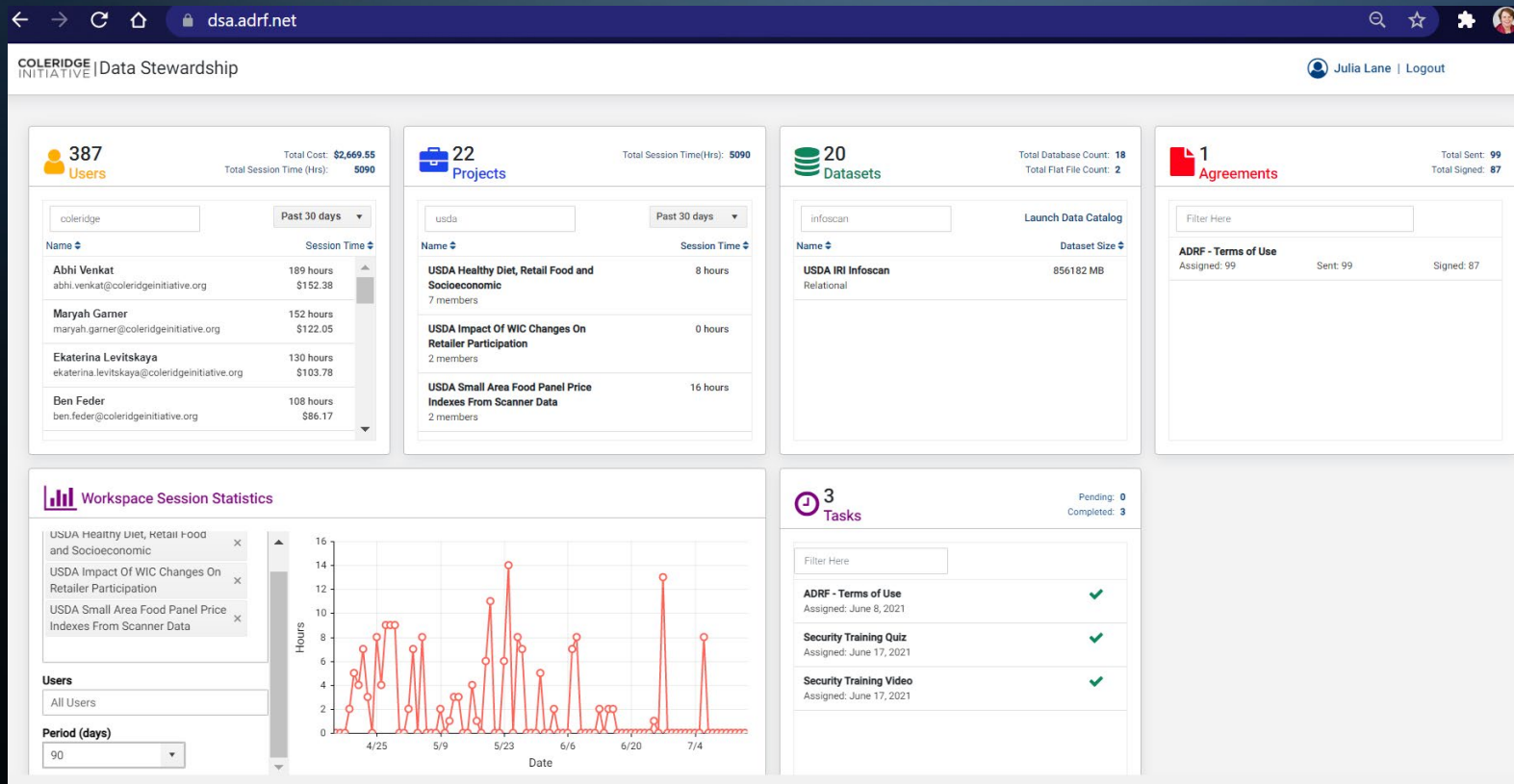
Security: Five Safes





Coleridge Initiative

Reporting: Data Stewardship





Approach

Agency

- Security: Five Safes
- Reporting: Data Stewardship
- Transparency: Pricing and Usage

Researcher

- Ease of use: Windows and Tools
- Customization: Choice
- Collaboration: Rich Context



Ease of Use

ADRF Getting Started

File | fs01/Software/GettingStarted/ADRF.html

Attention

Unauthorized access or use of this computer system may subject violators to criminal, civil, and/or administrative action. All information on this computer system may be intercepted, recorded, read, copied, and disclosed by and to authorized personnel for official purposes, including criminal investigations. Such information includes sensitive data encrypted to comply with confidentiality and privacy requirements. Access or use of this computer system by any person, whether authorized or unauthorized, constitutes consent to these terms. There is no right of privacy in this system.

Please keep in mind that any removal of data is prohibited. This includes taking screenshots or writing down numbers on paper. All output has to go through disclosure review. Violations of this policy may result in penalties and revoking access from the system.



Jupyter Notebooks

Learn the basics of Notebooks, accessing data with DBeaver & reading data using Python and R.



User Documentation

Access the Coleridge Initiative ADRF User Documentation



Support Link

For questions about using the ADRF or to request applications, libraries or resources please email us at support@coleridgeinitiative.org



Export Request

Initiate new or review the status of existing export requests



Data Catalog

Search and review the Data Catalog
Coming Soon!



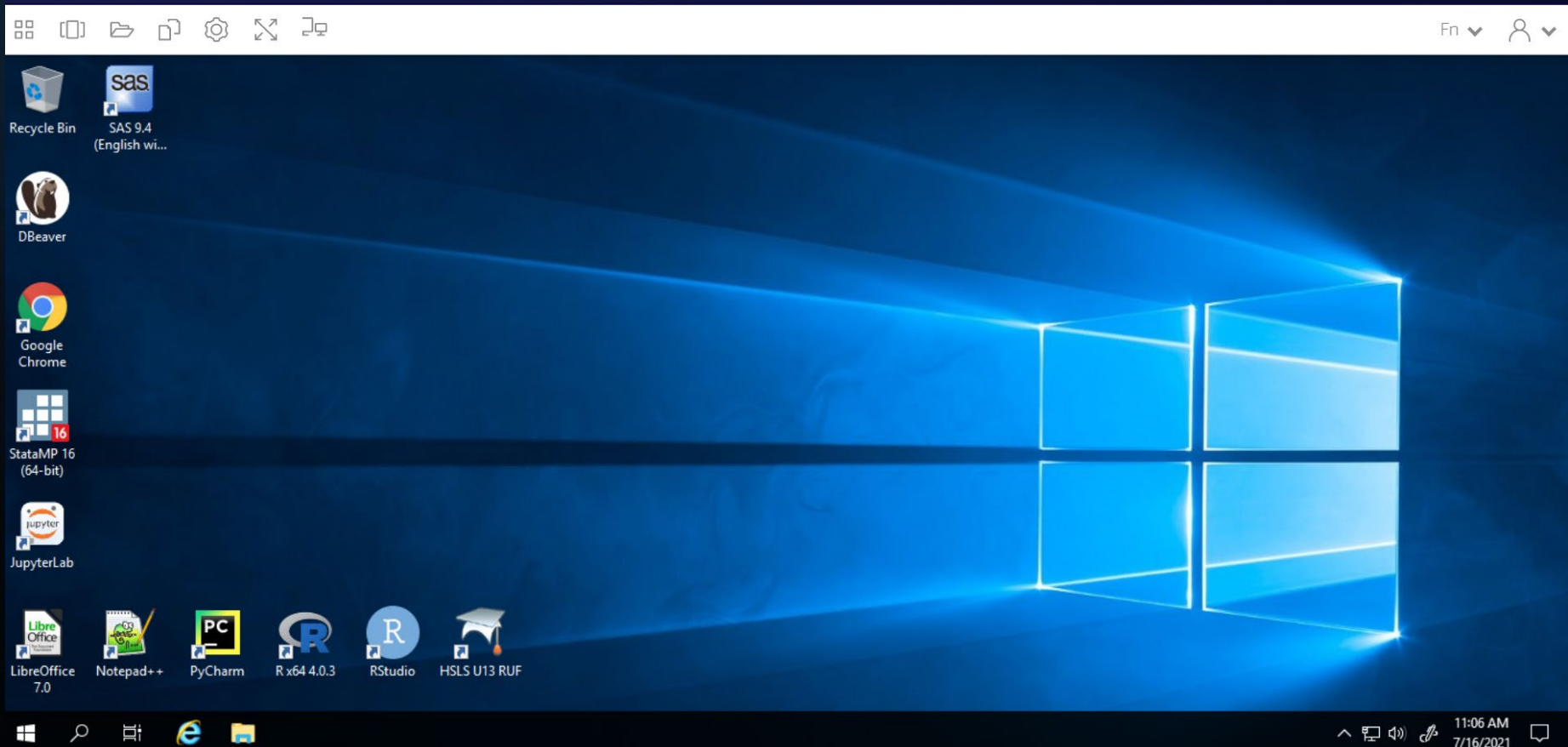
Usage Metrics

Review your usage of the ADRF platform
Coming Soon!



Coleridge Initiative

Ease of Use



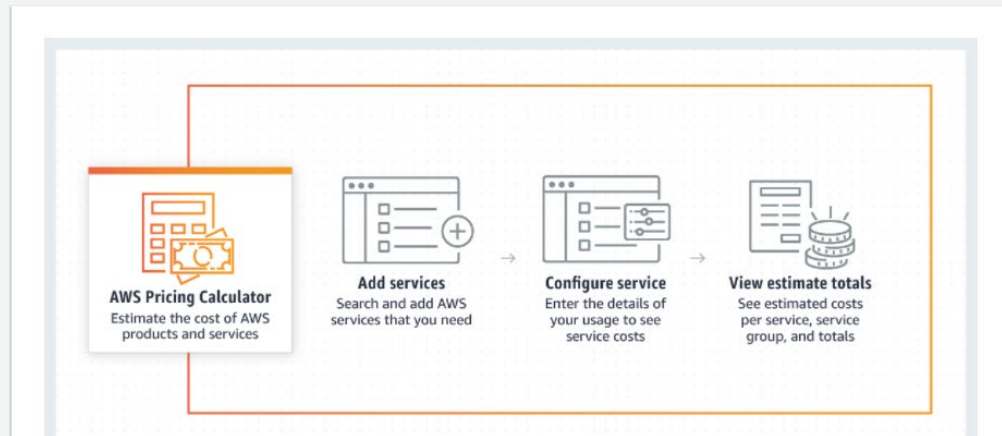


Customization

architecture solution.

Configure a cost estimate that fits your unique business or personal needs with AWS products and services.

How it works





Overview

- Context
- Secure Environment
- *Workforce Capacity*
- Search and Discovery
- Questions



Overview

- Context
- Secure Environment
- Workforce Capacity
- Search and Discovery
- Questions



Collaboration

Product Market Competition and Human Resource Practices in the Retail Food Sector

ELIZABETH DAVIS, MATTHEW FREEDMAN,
JULIA LANE, BRIAN MCCALL, NICOLE NESTORIAK,
and TIMOTHY PARK*

In the wake of Wal-Mart and other mass merchandisers' entry into food retailing, the nature of competition in the industry has changed radically. Using longitudinal data on workers and firms to construct measures of compensation and churning for traditional food retailers, this paper examines how these measures change in response to mass merchandiser entry. While there is considerable heterogeneity across retail food establishments, human resource practices are persistent even in the face of new external competition.

Introduction

Whereas Wal-Mart employees start at the same salary as unionized employees in similar lines of work, they make 25 percent less than their unionized counterparts

Data and Measurement

Constructing a data set that permits the analysis of the effects of changes in the product market on firm specific compensation policies requires information on firms, their workers, and the product markets in which they

⁴ This is different from the approach used by Neumark et al. (2005) since the research question is broader than examining the impact of the entry of one particular firm.

358 / ELIZABETH DAVIS ET AL.

operate. We rely on a new linked employer-employee database that also provides information on the location and industry of firms: the Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau (described in detail by Abowd et al. (2005) and Haltiwanger, Lane, and Spletzer (2006)). Briefly, the LEHD data consist of quarterly records of the employment and earnings of all workers who are covered by U.S. state unemployment insurance (UI) systems in the 1990s and early 2000s.⁵ About 96 percent of private wage and salary employment is covered by these



Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

Article

Authors

Metrics

Comments

Media Coverage



Abstract

Modeling the Framework
for False Positive
Findings

Bias

Testing by Several
Independent Teams

Abstract

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller: when effect sizes are

Why is This Important to Federal Agencies?

Evidence Act Title 2 – OPEN Government Data Act:

Section 202(c)

- Facilitate collaboration with non-Government entities (including businesses), researchers, and the public for the purpose of understanding how data users value and use government data
- Engage the public in using public data assets of the agency and encourage collaboration by publishing on the website of the agency, on a regular basis (not less than annually), information on the usage of such assets by non-Government users
- Assist the public in expanding the use of public data assets



Beginning in the Obama Administration, Agencies have been making datasets available for public use via Data. Gov. The Trump Administration augmented this by prioritizing data sets for AI R&D and those that support healthcare initiatives.

This has grown from a few datasets contributed by each Agency to today's status with over 300,000 datasets that are available in multiple formats, searchable, and tagged with industry protocols.

But just being available, does not mean that the data is "of value"

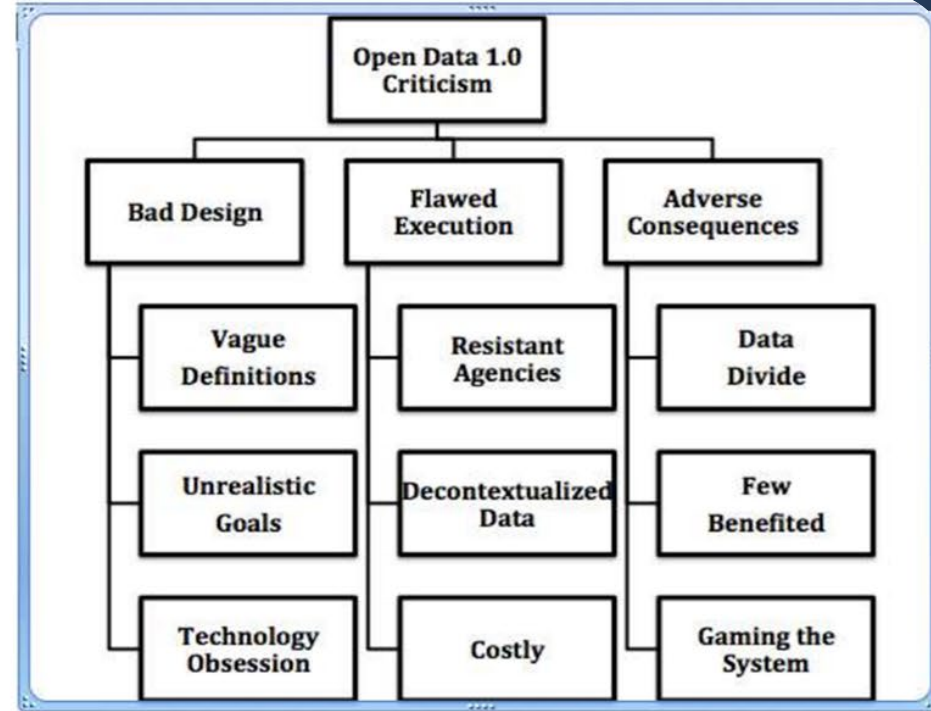


Figure 1: Open Data 1.0 Criticism

The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more](#).

For information regarding the Coronavirus/COVID-19, please visit [Coronavirus.gov](https://www.coronavirus.gov).

GET STARTED

SEARCH OVER 318,870 DATASETS





Economic Research Service
U.S. DEPARTMENT OF AGRICULTURE

[About ERS](#) | [Careers at ERS](#) | [FAQs](#) | [Contact Us](#)



[Home](#) | [Topics ▾](#) | [Data Products ▾](#) | [Publications ▾](#) | [Newsroom ▾](#) | [Calendar ▾](#) | [Amber Waves Magazine ▾](#)

[Home](#) > [Data Products](#) > [FoodAPS National Household Food Acquisition and Purchase Survey](#)

FoodAPS National Household Food Acquisition and Purchase Survey

[Overview](#)

[Background](#)

[Data Access](#)

[Summary Findings](#)

[Documentation](#)

[Interactive Charts](#)

[Research Projects and Publications](#)

[FAQs](#)

FoodAPS National Household Food Acquisition and Purchase Survey

USDA's National Household Food Acquisition and Purchase Survey (FoodAPS) is the first nationally representative survey of American households to collect unique and comprehensive data about household food purchases and acquisitions. Detailed information was collected about foods purchased or otherwise acquired for consumption at home and away from home, including foods acquired through food and nutrition assistance programs. The survey includes nationally representative data from 4,826 households, including Supplemental Nutrition Assistance Program (SNAP) households, low-income households not participating in SNAP, and higher income households. For a more detailed description of the survey, see [Background](#).

ERS has expanded access to FoodAPS data by removing identifying variables and publishing public-use data files (see below). Users can now download the public-use data in three file formats: SAS, STATA, and CSV. If the public-use files are not sufficient for the researchers' needs, see the [Data Access](#) page for instructions about how to gain access to restricted-use data.

The public-use files include the household-level and individual-level interview files, the food-at-home (FAH) and food-away-from-home (FAFH) event files, the FAH and FAFH item files, data from the Meals and Snacks form, a summary of the individual daily reporting status, household access to FAFH outlets and SNAP-authorized FAH retailers, and estimates of household SNAP eligibility. All codebooks and a User's Guide,

[Access SPARC resources related to COVID here](#) >

SPARC✖

Who We Are

What We Do

Why It Matters

Become a Member



TUESDAY, AUGUST 3, 2021 NEWS

On a Mission to Make Federal Data Sets More Useful and Accessible

Open Data

SHARE



Although government agencies manage massive amounts of information, there is little known about exactly how it is used and by whom: Turns out, there is little data on federal data.

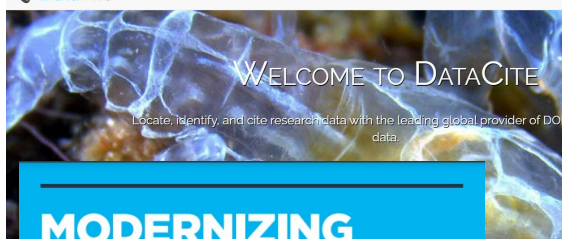


A new effort is underway to leverage artificial intelligence to better track what research is being done with what data. Ultimately, it could result in a data usage scorecard that could make it easier for researchers to find who else has used datasets in similar research, enabling them to reproduce results,

Create an ecosystem with incentives for Data as a first class asset

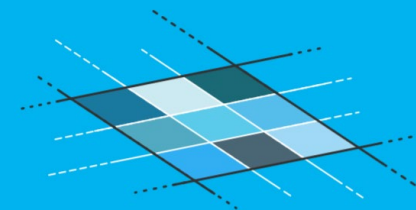
1. Find datasets in publications
 2. Document topics, authors, and ecosystem
 3. Document value of the dataset
- ...
- Agencies can show Congress (GAO), Agency Heads and the public; **get credit**
 - Researchers can find data experts and code; **get credit**
 - Publishers and editors can find reviewers; **get credit**
 - Taxpayers can see data use; **see value**

Plumbing is in place



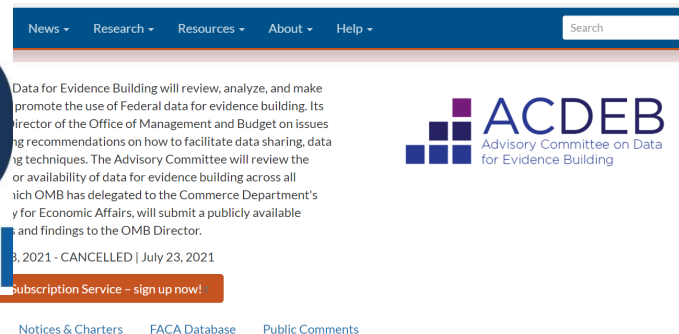
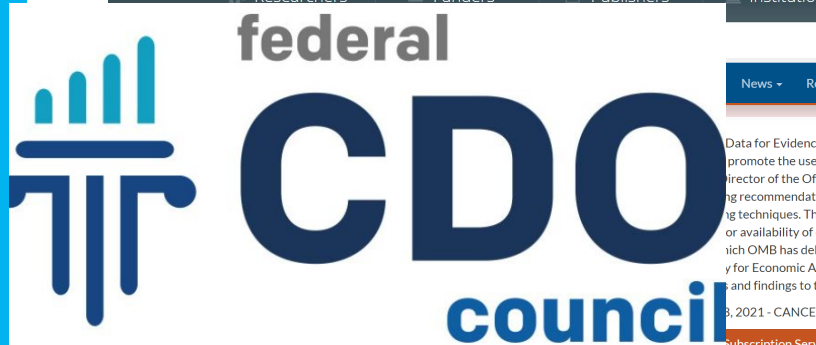
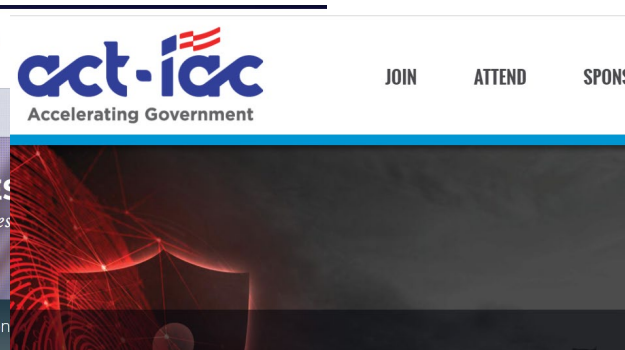
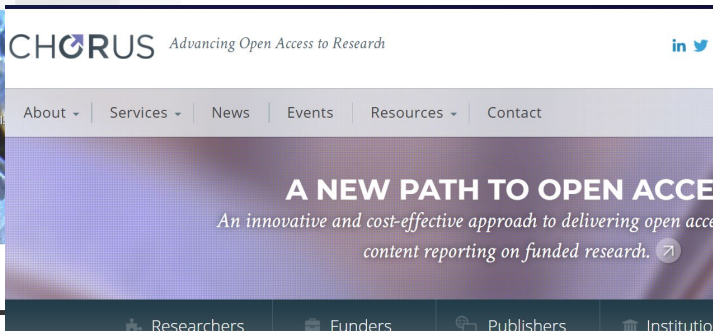
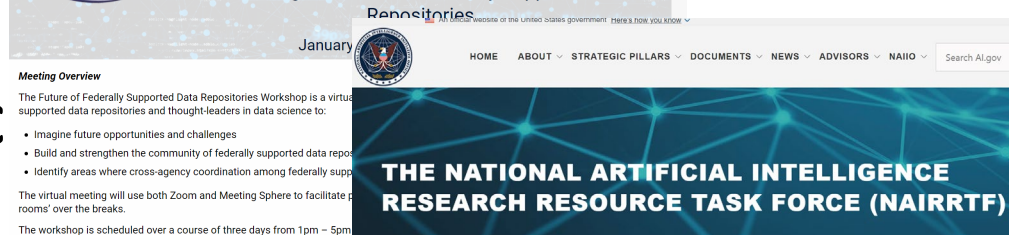
MODERNIZING U.S. DATA INFRASTRUCTURE:

Design Considerations for
Implementing a National Secure
Data Service to Improve Statistics
and Evidence Building



DATA FOUNDATION

Big Data Interagency Working Group Workshop:
Pioneering the Future of Federally Supported Data



How? Step 1: Find the data

Product Market Competition and Human Resource Practices in the Retail Food Sector

ELIZABETH DAVIS, MATTHEW FREEDMAN,
JULIA LANE, BRIAN MCCALL, NICOLE NESTORIAK,
and TIMOTHY PARK*

In the wake of Wal-Mart and other mass merchandisers' entry into food retailing, the nature of competition in the industry has changed radically. Using longitudinal data on workers and firms to construct measures of compensation and churning for traditional food retailers, this paper examines how these measures change in response to mass merchandiser entry. While there is considerable heterogeneity across retail food establishments, human resource practices are persistent even in the face of new external competition.

Introduction

Whereas Wal-Mart employees start at the same salary as unionized employees in similar lines of work, they make 25 percent less than their unionized counterparts

Data and Measurement

Constructing a data set that permits the analysis of the effects of changes in the product market on firm specific compensation policies requires information on firms, their workers, and the product markets in which they

⁴ This is different from the approach used by Neumark et al. (2005) since the research question is broader than examining the impact of the entry of one particular firm.

358 / ELIZABETH DAVIS ET AL.

operate. We rely on a new linked employer–employee database that also provides information on the location and industry of firms: the Longitudinal Employer–Household Dynamics (LEHD) Program at the U.S. Census Bureau (described in detail by Abowd et al. (2005) and Haltiwanger, Lane, and Spletzer (2006)). Briefly, the LEHD data consist of quarterly records of the employment and earnings of all workers who are covered by U.S. state unemployment insurance (UI) systems in the 1990s and early 2000s.⁵ About 96 percent of private wage and salary employment is covered by these



Search



Featured Code Competition

Coleridge Initiative - Show US the Data

Discover how data is used for the public good



Coleridge Initiative · 1,610 teams · a month ago

\$90,000

Prize Money

[Overview](#)[Data](#)[Code](#)[Discussion](#)[Leaderboard](#)[Rules](#)[Host](#)[Late Submission](#)

...

[Overview](#)[Edit](#)

Description

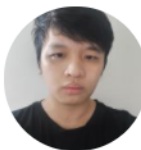
[Evaluation](#)[Timeline](#)[Prizes](#)[Code Requirements](#)[Coleridge Initiative](#)[Welcome](#)[+ Add Page](#)

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 agencies are now mandated to show how their data are being used. The new [Foundations of Evidence-based Policymaking Act](#) requires agencies to modernize their data management. New [Presidential Executive Orders](#) are pushing government agencies to make evidence-based decisions based on the best available data and science. And the government is working to respond in an [open and transparent way](#).

KAGGLE WINNERS



Nguyễn Tuấn Khôi - Zalo, VNG & Nguyễn Quán Anh Minh - Zalo AI, VNG

Context Similarity via Deep Metric Learning

Khôi and Minh are colleagues at VNG, their areas of expertise include natural language and speech processing, as well as deploying machine learning models to real world applications.

[Presentation Slides](#)

[Winning Model](#)

[Model Description](#)

[Interview Video](#)



Chung Ming Lee

Transformer-Enhanced Heuristic Search

Chun Ming is a Data Scientist active in the Singaporean startup scene. He's also worked as a Management Consultant at McKinsey & Co. and as a Software Developer in finance. Lee earned his MBA from London Business School and Bachelor's in Computer Science from Carnegie Mellon University.

[Presentation Slides](#)

[Winning Model](#)

[Model Description](#)

[Interview Video](#)



Mikhail Arkhipov

Pure Pattern Matching

Mikhail Arkhipov is from Moscow, Russia. Since 2017, he works on open-source NLP tools and performs research on Multilingual Transfer Learning and Named Entity Recognition.

[Presentation Slides](#)

[Winning Model](#)

[Model Description](#)

[Interview Video](#)

agency ▼	title ▼	training set labels ▼	ZALO_0_576 ▼	Chun_0_575 ▼	ZALO_0_565 ▼	Mikh_0_558 ▼	Osci_0_513 ▼	Kram_0_486 ▼	Car ▼
USDA	Economic Research Service: Specialized Agency Functions and Public Goodsâ€™ Provision	Agricultural Resource Management Survey FoodAPS National Household Food Acquisition and Purchase Survey	agricultural resources management survey arms national household food acquisition and purchase survey	agricultural resources management survey arms food acquisition and purchase survey foodaps	ers usda economic research service agricultural resources management survey foodaps arms national household food acquisition and purchase survey foodaps	arms agricultural resources management survey foodaps national household food acquisition and purchase survey	agricultural resources management survey arms	agricultural resources management survey national household food acquisition and purchase survey food	cor res pro ag res ma sur arn agr out
USDA	MEASURING WHAT WE SPEND	Agricultural Resource Management Survey FoodAPS National Household Food Acquisition and Purchase Survey National Postsecondary Student Aid Study	acs agricultural resource management survey american community survey brdis business r	agricultural resource management survey american community survey brdis business	national health and nutrition examination survey american community survey acs survey of	acs agricultural resource management survey american community survey bls consumer	acs agricultural resource management survey american community survey ces consumer	agricultural resource management survey national health interview survey national	cor exp sur ce u o sta bls

Coleridge Initiative - Rich Context API

API and SDK Documentation

Version: 0.0.3

This API represents the results of a joint pilot with NOAA, USDA, NSF, Elsevier, Chorus, and the Coleridge Initiative. The API provides an illustration of how agencies could access the metadata resulting from publication-dataset dyads produced by the "Show US the Data" project. We invite you to test this limited demonstration. You will need a token from Coleridge Support team to use the API. You will need to add a header with the token on each request.

You will need a token from Coleridge Support team to use the API. You will need to add a header with the token on each request.

<https://coleridgeinitiative.org>

Publication

getPublications

Get Publications from Rich Context Metadata Database

GET

/publication

Usage and SDK Samples

Curl Java Android Obj-C JavaScript C# PHP Perl Python

```
curl -X GET \
-H "api_key: [[apiKey]]" \
-H "Accept: application/json" \
"https://dtds1.coleridgeinitiative.org/api/v1.0.0/publication"
```

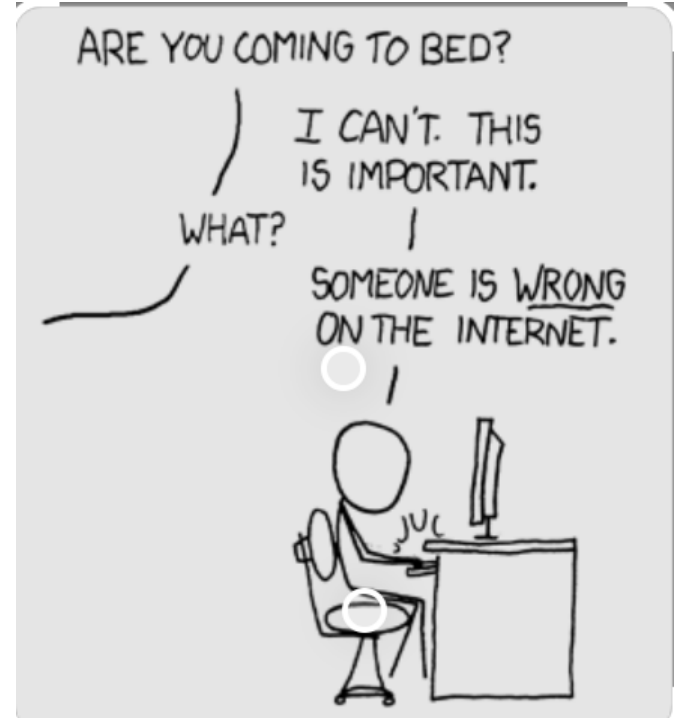
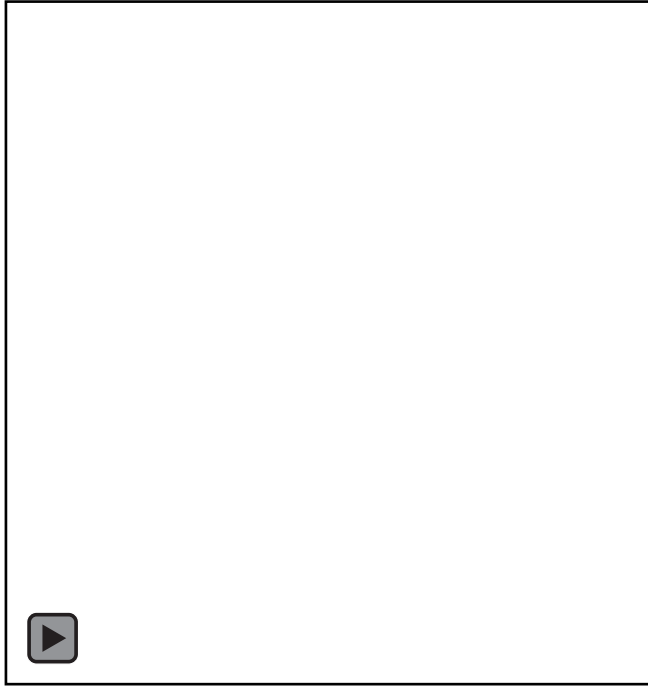

What can be done

[API documentation](#)

[TABLEAU dashboard](#)

[USDA Wordcloud](#) - Rural Urban continuum codes

Researcher community corrects errors



We Can Do It!



J. Howard Miller

POST FEB. 15 TO FEB. 28



WAR PRODUCTION CO-ORDINATING COMMITTEE

Show US the Data

October 20, 2021

Virtual

Register

ABOUT THE CONFERENCE

Evidence must be grounded in data and science. New legislation – the Foundations of Evidence-based Policymaking Act – requires agencies to report on the usage of their data, produce data inventories, and expand the public use of their data. The Office of Management and Budget requires data for evidence building across all agencies.

For subject expertise and to be successful, the new Federal Executive Order calls for data to be used to inform and guide decision-making. Data is the new gold.



Coleridge Initiative

Questions?

Please Contact:

Julia.lane@coleridgeinitiative.org

Julia.lane@nyu.edu