

# Theory of Change Showing Value of ERS Statistical Data

Spiro E. Stefanou  
Administrator, Economic Research Service  
Federal Committee on Statistical Methodology  
2023 Research and Policy Conference

October 24-26, 2023



# Data as a Product Innovation

## Specific Focus:

How are specific datasets being used in scientific and public research?

- Agricultural Resource Management Survey
- Rural Urban Continuum Codes

## Specific Application:

- Internal investment decisions
- USDA policy and strategies
- Congressional decision making

## Better Information For:

- Think tanks
- Researchers
- International agriculture statistics community



# Theory of Change Model

## Inputs

- Assessing the need for the data asset
- Survey development, piloting, execution
- Integrating survey data with proprietary data
- Curation

## Activities

- Natural language processing
- API
- Dashboards
- Notebooks

## Outputs

- Research and analysis
- Stronger communities
- Networks
- More information that can be used

## Outcomes

- Respondents respond more
- Policy makers and stakeholders have evidence-based insights
- Stimulate connections with interdependent systems

## Final Outcomes

- Congressional policy action
- Reassess scope of potential users
- Refine and innovate data assets
- Gain insights into emerging trends



# Assessing the Value of Public Data Assets

## Costs and Risks

- Acquisition, collection, curation, protection, storage
- Risk of disclosure, re-identification, & reputation to the agency

## Reward (or Utility) – more anecdotal

- Real value of these data to society, researchers and policy makers is yet to be determined
- Public provided data and information are special goods and offer a special challenge

## Project in Progress

- Develop an approach for evaluating the value of publicly available datasets and the potential value of free public access to these data
- Start with a proof of concept developing the basic methodology and then applying it to two distinct data sets
- Infometrics approach based on information theory

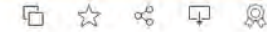






You will soon have access to your Tableau Public account using your Tableau login information. To learn more, please check out our [blog](#).

Democratizing Data - USDA by Democratizing Data



DATASETS  
3



PUBLICATIONS  
1,752



AUTHORS  
7,572



COUNTRIES  
58



CITATIONS  
14,626



INSTITUTIONS  
4,464

Select a Dataset to Explore Usage				Datasets: All, Year: 2017, 2018, 2019 and 3 more			CLEAR FILTERS
Name	Pub	Cit		1,752 Publications		DOWNLOAD SPREADSHEET	
RUCC	1,033	10,335		<b>Publication</b>	Pub	Cit	
NASS Census of Agriculture	666	3,798		CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015	968		
Agricultural Resource Managem..	89	632		CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2..	793		
				Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated?	712		
				Survival after minimally invasive radical hysterectomy for early-stage cervical cancer	325		
				CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2013-2017	264		
				Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: A population-based study	223		
				Rural-Urban differences in cancer incidence and trends in the United States	143		
				Brain metastases in newly diagnosed breast cancer: A population-based study	134		
				910 Journals		DOWNLOAD SPREADSHEET	
				<b>Publication Title</b>	Pub	Cit	
				Journal of Rural Health	52	405	
				Journal of Soil and Water Conservation	38	114	
				International Journal of Environmental Research and Public Health	32	145	
				Sustainability (Switzerland)	30	171	
				Applied Economic Perspectives and Policy	24	191	
				4,464 Institutions		DOWNLOAD SPREADSHEET	
				<b>Institution Name</b>	Pub	Cit	
				RAND Corporation	19	188	
				Department of Agricultural and Resource Economics, Colorado State University	14	41	
				College of Nursing, University of Kentucky	11	35	
				Department of Agricultural and Resource Economics, University of Tennessee	11	18	
				University of North Carolina at Chapel Hill	11	86	
				Department of Agricultural Economics, Purdue University	10	78	
				Department of Agricultural Economics, Kansas State University	9	30	
				Department of Sociology, Iowa State University	8	84	
				Holden Comprehensive Cancer Center, University of Iowa	8	9	
				Department of Agricultural Economics and Economics, Montana State University	7	7	

Filter by Year(s)

Year	Pub	Cit	Authors
2017	169	2,801	667
2018	256	4,146	1,037
2019	320	3,693	1,427
2020	359	2,862	1,557
2021	505	1,100	3,515
2022	143	24	678

Filter by Topic(s)

jupyter Latest\_Basic\_information (1.25 1830) (1) Last Checkpoint: a minute ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted | P

Run Code

### Retrieve all agency runs

The queries below search for data for a specific ML run performed by Elsevier. All the different runs are stored in the `agency_run` table. Here we

```
In [67]: sql="sel
agency_r
agency_r
```

```
Out[67]:
```

agency
rucc

### What topics are an agency's datasets being used to study?

```
In [64]: sql=f"""
with a as (
select ds.id as ds_id, max(ds.alias) as dataset
, t.id as topic_id, max(t.keywords) as topic
, count(distinct p.id) as pub_per_year
, rank() over (
from agen
join dyac
join publi
join publi
join topi
join data
join data
```

### 1) Basic Usage Information

How much are agency datasets used in research and how has that usage changed over time? How often is each one of an agency's identified dataset used in research and how has that usage changed over time?

### How much are agency datasets used in research? or How often is each one of an agency's identified dataset used in research?

The table shows the use of each dataset based on publications that have used respective dataset.

```
In [61]: sql=f"""
select ds.alias as dataset, p.year
, count(distinct d.publication_id) as pub_per_year
-- using a python f-string so that parameters AGENCY and VERSION
-- count the different publications
```



# Democratizing the Data

## Webinar – March 28

- ERS & NASS approach to understanding how their data are used that provides a more in-depth look that can help inform decision-making around their data investments and improve the quality of data that are collected.

## Workshop May 1, 2023 (in-person)

- Goal: Gather community input into the usage statistics being developed in response to the [Evidence-based Policymaking Act](#).
- Participants had the opportunity to network with other users, as well as work hands-on in small groups to provide feedback on the new tools that USDA has developed – [usage dashboards](#), [Jupyter Notebooks](#) & an [API](#).
- Using the information about authors, institutions, and topics to find related literature and other experts, participants could discover how those datasets are being used in scientific and public research.





# Data Users Workshop: Early Lessons

## 1. Expansion

- Include more data
- Combine new data sources (GIS) with USDA data
- Expand corpus of publications beyond scholarly

## 2. Engage the Community to Improve Data Quality

- Gamify identifying parent datasets & provide fingerprint as part of the data system
- Input on traceability/lineage/fingerprinting of data
- Develop human-computer interaction to update the corpus
- Engage MSIs through target workshops and dashboard updates





# Data Users Workshop: Early Lessons

## 1. Develop and Deliver Training on Building Dashboard

“How to build your own dashboard”

## 2. Engage in Proactive Marketing



# Advancing DEIA with Data Democratization

## Inclusion and Accessibility:

- Lowering the cost to finding networks
- MSI Scholars
- Stakeholders
- Data providers
- State and local Governments
- Extension
- Commodity groups



# Expanding Data Accessibility

## Advancing Access to ERS Data (ERS and NYU)

- Goal: engage with external partners to understand how ERS data are used, constraints with using the data, and ways to encourage greater access across diverse stakeholder groups.
  - Extension Foundation,
  - state agencies,
  - academic institutions, and
  - other organizations
- ERS and NYU will engage in a set of intentional activities with specific organizations, including Minority Serving Institutions (MSIs) to determine how to support and expand their use of USDA statistical data.

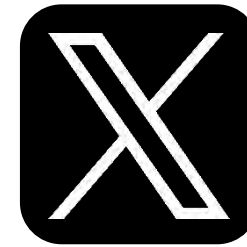
# Like, Share, & Follow ERS



[www.ers.usda.gov](http://www.ers.usda.gov)



[www.ers.usda.gov/  
data-products/charts-of-note](http://www.ers.usda.gov/data-products/charts-of-note)



@USDA\_ERS



[linkedin.com/company/  
usda-economic-research-service](https://linkedin.com/company/usda-economic-research-service)

**Subscribe to Weekly E-mail Notifications:** [www.ers.usda.gov/subscribe](http://www.ers.usda.gov/subscribe)

**Learn About Careers at ERS:** [www.ers.usda.gov/about-ers/careers-at-ers](http://www.ers.usda.gov/about-ers/careers-at-ers)

