Envisioning food and agricultural data for the future:

Engaging communities to identify trusted data

Lauren Chenarides, Ph.D.

Assistant Professor, Colorado State University

C-FARE Board Member

Farmers Sue Agriculture Department Over 'Purge' of Climate Information

Published Feb 24, 2025 at 6:50 PM EST | Updated Feb 24, 2025 at 7:05 PM EST

How do you report on the weather when data is disappearing?

"The U.S. has a system that is the best in the world, but we're actively sabotaging it."

By NEEL DHANESHA @neeldhanesha.com May 19, 2025, 1:02 p.m.

U.S. data quality has been declining for years. Now Trump's cutbacks are leading economists to question its figures

BY MARCO QUIROZ-GUTIERREZ
REPORTER

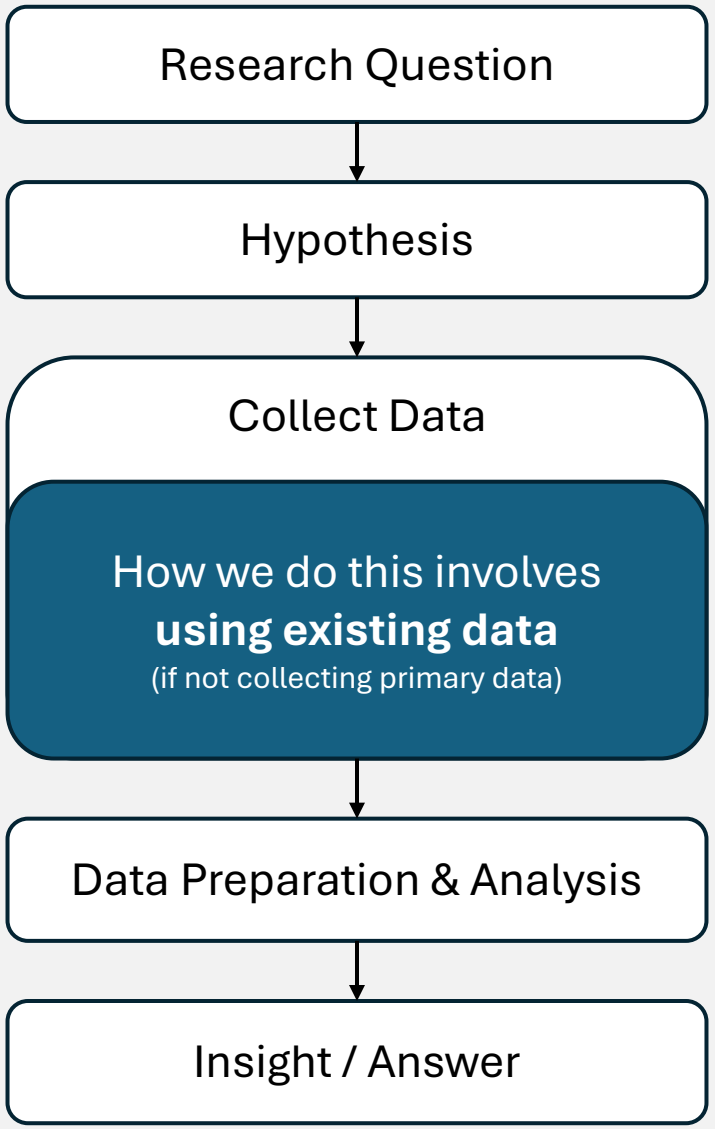
June 13, 2025 at 12:37 PM EDT

Economists Raise Questions About Quality of U.S. Inflation Data

Labor Department says staffing shortages reduced its ability to conduct its massive monthly survey

By Matt Grossman Follow

Updated June 4, 2025 3:27 pm ET



According to an informal questionnaire, webinar registrants noted data access disruptions:

- Disrupted access to **Census**, e.g., the **Community Resilience** dataset, **USAID DDL**, and **DEC**
- Increased difficulty navigating public health data, including **CDC BRFSS** and **YRBSS**
- Discontinued access to **BEA Regional Datab, EPA** datasets like the **Environmental Justice Screener**
- Researchers unsure about continued access to **USDA and EPA datasets**

Many of us depend on publicly available data to answer food and agricultural research (FAR) questions

We s
corpu
of 12
found
these
agric

Results are
https://dem
research/#

ARMS)
plement

rough USDA TPAA)

ural land (TOTAL)

Chenarides, L., Bryan, C., & Ladislau, R. (2025). Methodology for comparing citation database coverage of dataset usage. Available at: https://laurenchenarides.github.io/compare_scopus_openalex_report/report.html

Yet, researchers are looking *beyond federal data* to answer FAR questions.

Examples of non-federal data used in food economics research:

What is the impact of **SNAP work requirements** on household food spending and **diet quality**? (Zhang and Fitzpatrick, 2024)

How can mobility patterns help identify seasonal or geographic vulnerabilities to **food insecurity**? (Zufiria et al., 2018)

How do **food insecurity** and **nutrition insecurity** differ, and how are they associated with health outcomes? (Livings et al., 2022)

How can we measure the **healthiness** of **grocery purchases** over time and across populations using commercial product data? (Lin et al., 2025)

Our analysis of Dimensions data also suggests that **alternative sources may be widely used.**

Across publications in the applied economics & agricultural sciences fields:

Only 1.75% of **food insecurity** studies use one of the 12 publicly funded datasets.

Only 8.6% of **food pantries** studies reference them.

How do we find what alternative datasets are being used?

Mentions of datasets can be found in publications by applying ML models and other methods.

Identify dataset mentions by scanning research papers for sentences that refer to datasets, such as “we used data from…” or “data were obtained from…”.

Preprocess the text by standardizing and cleaning the training examples by removing inconsistencies, simplifying formatting, and correcting incomplete annotations.

Evaluate results based on how accurate the system is (correctness) and how many relevant mentions it captures (coverage).

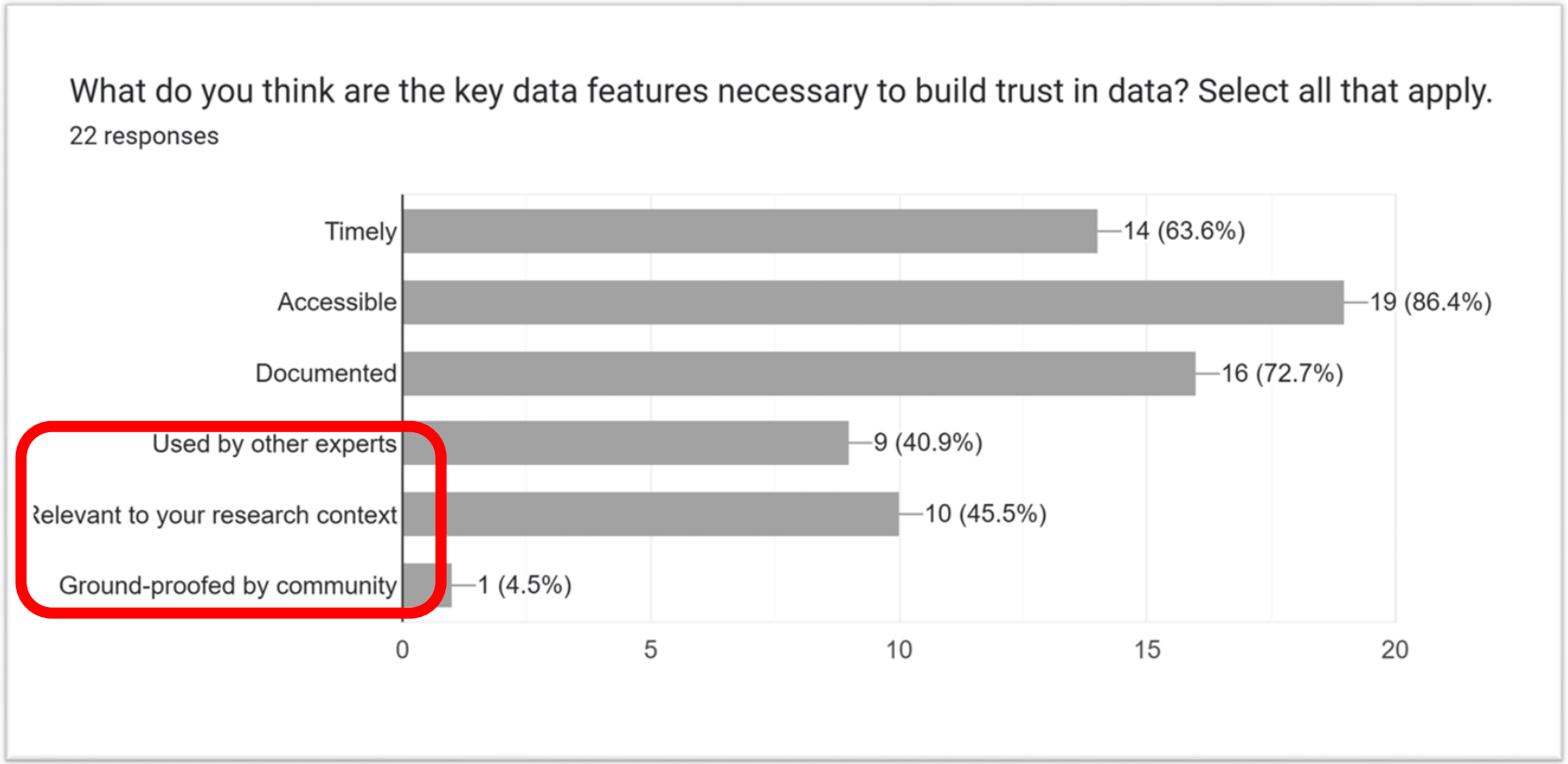
Harvard Data Science Review 2024 Democratizing Data: Discovering Data Use and Value for Research and Policy https://hdsr.mitpress.mit.edu/specialissue4

Hausen, Ryan, and Hosein Azarbonyad. "Discovering data sets through machine learning: An ensemble approach to uncovering the prevalence of government-funded data sets." *Harvard Data Science Review* Special Issue 4 (2024). https://hdsr.mitpress.mit.edu/pub/ou89oggk/release/7?readingCollection=a6db7dff

Chenarides, L., Bryan, C., & Ladislau, R. (2025). Methodology for comparing citation database coverage of dataset usage. Available at: https://laurenchenarides.github.io/compare_scopus_openalex_report/report.html

When we find the alternative datasets, which ones do we trust?

We polled webinar registrants about what it means for data to be **trusted**.



Possible elements to be checked from alternative data sources to establish trust:

General fundamentals

- ❑ Technical integrity
- ❑ Documentation
- ❑ Relevance and coverage
- ❑ Usability and accessibility
- ❑ …

Trust is likely to be context-specific

- ❑ Data are appropriate for the decision or question being asked
- ❑ Relevant to the user's setting, timeframe, and scale
- ❑ Interpretable and actionable for the intended audience
- ❑ Aligns with real-world use cases
- ❑ …

Expert communities will be necessary to validate the fundamentals and certify the fit for purpose.

Spector, Alfred Z., Peter Norvig, Chris Wiggins, and Jeannette M. Wing. "Data science in context: Foundations, challenges, opportunities." (2022).

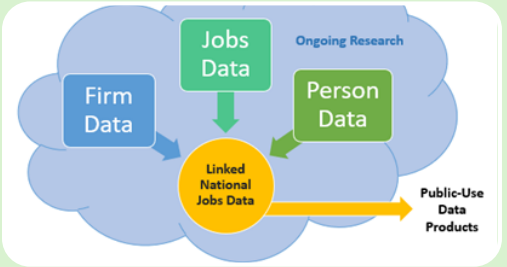
Examples of community-driven models built around a central question:

Technology & Innovation

What can we learn about patterns of innovation by linking information on inventors, organizations, and technologies across U.S. patents?

Research & Development

What is the impact of university research spending on workforce development, regional economies, and career pathways?



Labor

How can we link workers to employers over time to better understand employment patterns, earnings trajectories, and local labor markets?

What would a community-driven model look like for **central questions** in **food & agriculture**?

To begin answering key questions, we have three excellent presenters:

1. **Nick Pallotta (USDA NASS)** – How can we build ag community networks so that we can learn about data quality?
2. **Mark Locklear (Extension Foundation)** – How can we provide real-time insights that are immediately useful?
3. **Julia Lane (NYU)** – What new data infrastructures are on the horizon that could build and support communities?

As you're listening, ask yourself:

- How do we build communities to identify alternative data sources?
- What are the central FAR questions that would benefit from a community-driven approach?