

Report on the
USDA Data Usage Workshop
Hosted on June 10, 2024
ABRIDGED VERSION

Prepared by:

Lauren Chenarides, Ph.D.
Assistant Professor
Colorado State University

Reviewed by:
USDA-NASS POC

Last updated:

July 17, 2024

Table of Contents

1. Executive Summary -----	1
2. Introduction -----	2
2.1 Background of the Dashboard -----	2
2.2 Purpose of the Workshop -----	2
3. Summary of Activities -----	3
3.1 Workshop Structure and Activities -----	3
3.2 Activity 1: The Ideal Tool -----	4
3.3 Quick Poll Results -----	5
3.4 Activity 2: Dashboard Demo and Walkthrough -----	6
3.5 Large Group Share: Summary of Reflection Question Responses -----	8
4. Dashboard Highlights -----	10
4.1 Gaps Filled by the Dashboard -----	10
4.2 Value to Different User Communities -----	11
5. Suggestions for Improvement -----	12
5.1 Short-term Improvements (Minimal to Low Investment) -----	12
5.2 Long-term Developments (Larger Investment/Funding Required) -----	13
6. Expanding User Engagement -----	13
6.1 Reaching Out to Different User Communities -----	13
6.2 Defining Valuable Usage Metrics for Different User Communities -----	14
6.2 Potential Future Collaborations -----	15
7. Conclusion and Next Steps -----	15
Appendices -----	A
A1. Method for Report Writing -----	A
A2. Workshop Details -----	B
A3. IdeaBoardz -----	B
A4. Breakout Session Transcripts -----	B
A5. Participant List -----	B

1. Executive Summary

On June 10, 2024, a workshop was convened to gather feedback on the DemocratizingData.ai dashboard, a tool developed by the United States Department of Agriculture (USDA) and its partners to track the usage of USDA-produced datasets. This initiative, inspired by the Foundations for Evidence-Based Policymaking Act of 2018, aims to provide data usage statistics about how public data are used across various sectors.

The workshop brought together 34 participants, primarily research economists who are heavy users of the datasets currently featured on the dashboard. The workshop was hosted by a team of researchers in the Department of Agricultural and Resource Economics at Colorado State University, each of whom are also heavy data users of USDA-produced datasets. Through a series of structured activities, the workshop sought to understand how effectively the dashboard meets the needs of the research community and to identify areas for improvement and expansion.

Key findings from the workshop include:

- The dashboard fills a significant gap as a centralized hub for information on dataset usage, and was valued by participants. Features such as identifying dataset-specific experts and providing direct links to research articles were deemed particularly useful, with over 75% of respondents rating these features as very or somewhat valuable.
- Participants emphasized the need for more comprehensive and accessible documentation for datasets, improved search and filter functionality, and information on dataset linkability. These insights highlight opportunities for both short-term improvements and long-term developments of the dashboard.
- While the audience of the workshop was research economists, the workshop revealed potential value for a broader range of stakeholders, including early career researchers, policymakers, and data contributors. This suggests opportunities for expanding the dashboard's reach and impact.
- The workshop also identified challenges in tracking dataset usage and defining valuable usage metrics for different user communities. To maximize the dashboard's utility and ensure its relevance across diverse user groups, some suggestions that were made during the workshop were creating metrics that resonate with different user communities.

The DemocratizingData.ai dashboard represents a significant step toward making USDA data more accessible and expanding their reach. By building on the insights gained from this workshop, the dashboard has the potential to become a sought-after tool for data-driven research and evidence-based policymaking.

The feedback and recommendations from this workshop were conducted by evaluating the transcripts of the breakout rooms, notes from the note takers, and additional data collected from the activities and the quick poll. A detailed overview of the methods for writing this report is available in [Appendix A1](#). As the project moves forward, it is recommended to continue to evaluate the utility of the dashboard by holding workshops for various user communities to gauge that the dashboard aligns with their needs.

2. Introduction

2.1 Background of the Dashboard

The DemocratizingData.ai dashboard is a digital platform initially developed with funding from Schmidt Sciences, Alfred P. Sloan Foundation, Overdeck Family Foundation, and Patrick J. McGovern Foundation. It was then further developed by USDA ERS, NASS, NCES, and NCSES. The Democratizing Data initiative harvests knowledge from millions of scientific publications and presents information about who is using the data, for what, and how the data have been used in its usage dashboards. The dashboard represents a significant step toward increasing transparency, accessibility, and utility of USDA data across various user groups and disciplines.

The dashboard was created as part of a broader effort to democratize access to data and promote evidence-based policymaking in response to the Foundations for Evidence-Based Policymaking Act of 2018. The initiative aims to provide a centralized platform where a diverse set of users - including researchers, policymakers, stakeholders, and the general public - can easily identify USDA-produced datasets, understand their usage patterns, and gauge their impact. By leveraging machine learning and data visualization techniques, the dashboard offers insights into dataset usage, publication trends, and the broader influence of USDA data across various fields and applications.

2.2 Purpose of the Workshop

To gather insights on the DemocratizingData.ai dashboard, a workshop was convened on June 10, 2024. The participants were a subset of the potential user community, specifically research economists who are heavy users of the datasets currently featured on the dashboard (see [Table 1](#)). While this group represents a portion of the target user base, it is important to note that they do not encompass the full range of potential users for the dashboard.

The primary objectives of this workshop were:

1. To gather insights from experienced USDA-data users, gauging how effectively the dashboard meets the needs of the research community.
2. To collect feedback on the dashboard's features, usability, and overall value to potential users.
3. To identify gaps in current dataset usage that the dashboard addresses, as well as opportunities for enhancement.

The workshop was structured to encourage active participation and candid feedback through a series of guided activities and discussions.

This report summarizes the key findings, insights, and recommendations that emerged from the workshop. While the insights gathered are primarily from the perspective of research economists,

they are meant to be informative in shaping a tool that can enhance the overall accessibility and utility of USDA datasets for researchers, policymakers, and other potential stakeholders.

Table 1. List of Datasets Featured on Dashboard

Dataset Name	Breakout Group*
NASS Census of Agriculture	Group01
Agricultural Resource Management Survey (ARMS)	Group01
Food Acquisition and Purchase Survey	Group02
Current Population Survey Food Security Supplement	Group03
Food Access Research Atlas	Group04
RUCC	
Household Food Security Survey Module	
Local Food Marketing Practices Survey	
Farm to School Census	
Quarterly Food at Home Price Database	
Tenure Ownership and Transition of Agricultural Land	
Information Resources Inc. (IRI) InfoScan	
Transition of Agricultural Land Survey	

*Group 1 was assigned two datasets because these datasets are often used together by the research economist community of participants.

3. Summary of Activities

The workshop was designed to gather feedback on the DemocratizingData.ai dashboard through a series of structured activities. Thirty-six economists accepted our invitation to attend the workshop. Two did not attend. The invited participants consisted of research economists who are heavy users of the datasets currently featured on the dashboard. This group represented a subset of the potential user community, chosen for their expertise and familiarity with USDA datasets. A full list of participants is available in [Appendix A5](#).

3.1 Workshop Structure and Activities

The workshop was structured around breakout sessions allowing for more interactive discussions. The participants were divided into four breakout groups (see [Table 1](#)), each with up to 10 individuals, with breakout-room assignments based on research areas or sub-discipline specialties that aligned with each discussant's expertise. This approach ensured that each group had a mix of expertise relevant to specific aspects of the USDA datasets. A full agenda of the workshop is available in [Appendix A2](#).

Each breakout group was assigned a discussant to facilitate conversation and a note taker to record key points and insights:

- Group 1: Allie Bauman (Discussant), Brooke Fitzgerald (Note Taker)
- Group 2: Becky Cleary (Discussant), Taís de Menezes (Note Taker)
- Group 3: Alessandro Bonanno (Discussant), Mackenzie Gill (Note Taker)
- Group 4: Lauren Chenarides (Discussant), Sione Ikeme (Note Taker)

The workshop was moderated by Mayla Boguslav, who kept track of time and ensured smooth transitions between activities.

The workshop was comprised of two main activities conducted within the breakout rooms:

1. [Activity 1](#): Conceptualizing an ideal tool for tracking dataset usage statistics. This activity used IdeaBoardz, an online collaboration tool, to gather participants' thoughts on key pieces of information, tools, and features of a tool that collects data usage statistics.
2. [Activity 2](#): Exploring and providing feedback on the current dashboard prototype. Participants were given a guided tour of the dashboard and then asked to interact with it, focusing on specific datasets assigned to their group.

Following these activities, all participants reconvened to share their experiences and insights from the breakout sessions. This final discussion allowed for participants to share their feedback collectively.

3.2 Activity 1: The Ideal Tool

In the first activity, participants were asked to conceptualize an ideal tool for tracking dataset usage statistics. They provided input on three key questions: (1) What key pieces of information would you want to know about the usage of this dataset?, (2) Are there tools / resources that can provide you with the information you want?, and (3) What features does, or should, a tool / resource have to help you decide if the dataset is useful? Responses from each group are available in [Appendix A3](#).

Here are the main findings from each group:

Group 1:

- (1) Key pieces of information: ease of accessing full data for research, data documentation and variable definitions, and frequency/granularity of observations.

- (2) Tools/resources: publicly available data documentation, published papers, and USDA reports using the dataset.
- (3) Features: sample code in popular languages for data loading and cleaning, crosswalks to other datasets, and user-friendly formats.

Group 2:

- (1) Key pieces of information: variables and their definitions, journals where previous authors have published, and intended users of the data.
- (2) Tools/resources: journal article databases like PubMed and Scopus, and detailed codebooks from data owners.
- (3) Features: standard documentation across datasets, clear information on public vs. restricted-access data, and a way to track datasets across journals.

Group 3:

- (1) Key pieces of information: available years of data, who is using the data and where, and data granularity across time and geography.
- (2) Tools/resources: replicable examples or descriptive statistics, recommendations on using weights and complex survey adjustments, and clear documentation.
- (3) Features: information on the length of the application process duration for accessing restricted use data, and lists of working papers and authors using the data.

Group 4:

- (1) Key pieces of information: data lag, update frequency, and geographic units available.
- (2) Tools/resources: working paper series showing data usage across disciplines (similar to the SSRN page for Kilts data users), and direct access to data dictionaries and documentation.
- (3) Features: ability to download data in analysis-ready formats, and query functions to find authors or journals using specific datasets.

3.3 Quick Poll Results

Following the ideation activity, participants were polled on two features of the existing dashboard before they interacted with it. These questions were meant to gauge the perceived value of two existing features of the usage statistics available on the dashboard. Out of 34 workshop participants, 27 responded to the poll, representing a response rate of approximately 79%. The poll used a 5-point scale with the following key:

- 1 - Not Valuable: This feature would not influence my decision at all.
- 2 - Slightly Valuable: This feature has minimal impact on my decision.
- 3 - Neutral: This feature would not strongly influence my decision either way.
- 4 - Somewhat Valuable: This feature would be useful but not decisive.
- 5 - Very Valuable: This feature would significantly influence my decision.

The results for each proposed feature were as follows:

1. Consider a feature that has direct links to research articles using this dataset. How valuable would this feature be in making your decision?
 - 67% of respondents (18 out of 27) rated this feature as very valuable (5/5)
 - 22% (6 respondents) rated it as somewhat valuable (4/5)
 - 7% (2 respondents) were neutral (3/5)
 - 4% (1 respondent) rated it as slightly valuable (2/5)
 - No respondents rated it as not valuable
 - The average rating was 4.51/5, suggesting this feature has high perceived value
2. Consider a feature that identifies dataset-specific experts and their institutions. How valuable would this feature be in making your decision?
 - 30% of respondents (8 out of 27) rated this feature as very valuable (5/5)
 - 44% (12 respondents) rated it as somewhat valuable (4/5)
 - 22% (6 respondents) were neutral (3/5)
 - 4% (1 respondent) rated it as slightly valuable (2/5)
 - No respondents rated it as not valuable
 - The average rating was 4.0/5, indicating high perceived value

These quick poll results indicate that both features are valued by the participants who responded, with a stronger preference for direct links to research articles.

The raw data used to summarize the results of the quick poll are not included in this report but are available upon request.

3.4 Activity 2: Dashboard Demo and Walkthrough

During the second activity, participants were given a demonstration of the dashboard and asked to explore it themselves, focusing on specific datasets assigned to each group. The groups provided feedback on various aspects of the dashboard within their breakout sessions. Excerpts from the discussions are included below, and the full transcripts are available in [Appendix A4](#).

Group 1:

- “This premise of thinking about organizing publications through the data that they’re using is one that is not common. I’m not sure I’ve experienced it before in economics. So it is interesting to think about searching through the literature through datasets. Obviously, I think we’re identifying a lot of the road bumps, but the premise to me is interesting and new.” (Assistant professor)
- “So my comments were mostly about just functionality when I was searching for things and placement, mostly eye placement on the page so that you find the information that you need faster. So having author at the top because that’s most likely what we’re going to be searching for. Then also, some of the placement within the search menu as well. Having the option to hide that so that we can better see the screen when we’re actually looking through the information. Then also wondering what is the end goal in terms of

how this information is going to be used by the researchers.” (Agricultural economist at ERS)

- “Ag Census is very broad. So if there was a way to limit that or break that down a little bit, so is it microdata? Is it county level data? Is it just current 2017, 2022 data or is it historical all the way back to 1850? Some of that information would be really helpful for some of the work that I’ve been doing. So it was just very broad and there’s just so much, I think there’s 431 publications with Ag Census and they’re just all over the place, so just being able to narrow that down would be great.” (Agricultural economist at ERS)

Group 2:

- “I think it’s an exciting tool. I think that there’s a lot of really cool things. I do think there’s a trade-off with the more things that you include. It gets harder to sort your way through it. And I think that it’s in kind of a good area in terms of being somewhat streamlined, but having enough information. I like the amount of things that you can click on and modify by. I think that’s cool to see the filtering. I think that that’s interesting.” (Agricultural economist at ERS)
- “I think that being able to track publications and authors and getting connected to who’s using those data would be really helpful for me and for our agency.” (Branch chief at ERS)
- “It looks like it’s still kind of the same now, is that a lot of the identified publications don’t use the data, which I do think that that’s something that’s still happening, where a lot of them, I opened three or four that I know... I have looked at that don’t use the data where a lot of times, it’s a title in the citation section. And so I don’t know if that’s something that you guys have thought about at all, how to try to minimize that?” (Agricultural economist at ERS)
- “I think a word cloud related to the methods or estimation techniques would also be useful. Because then keyword, but then you can filter on both and you’d be like, oh, there is causal availability here, or it’s just descriptive, or those types of things.” (Assistant professor)

Group 3:

- “I think that the key things that are featured here are the same ones I would have prioritized. How do you find the publications? How do you find the journals? How do you find the authors that are publishing in an area?” (Full professor)
- “In probably most of our cases, is for us, we already would know data set specific experts, and we would know direct links to reach, because we we’re well involved in these literatures. Now, for somebody started out in a third-year in a PhD program, this could be very useful information. So, I want to figure out how you want us to respond to this. For ourselves, or for a third-year PhD student? ... I think this is a great tool. I think it's probably a lot more relevant... I think you should try to figure out which group you want to tailor this towards. It seems like it may be more useful for graduate students, or more recent PhDs. And so, I think, for a lot of us that are in this call, it may be less relevant for us, as we're farther along in our careers. And so, I think by designing this and

figuring out which group is it most useful for. I think it's great for graduate students or fresh PhDs.” (Full professor)

- “Just with this particular data set, because there's so many names that it goes by, you could say December supplement. You could say also... I mean, it's almost like the March supplement, outgoing rotation group, and all that. So, this one might be a unique problem for the search engine. You can relay that to them, that there's multiple names for this data set. Unlike food apps, there's only one. I guess maybe there's more. I don't know.” (Associate professor)

Group 4:

- “Some people may have issues looking at certain colors. I know that with our mapping tools, we have to make sure that we meet those things just to make sure it's in readable format for everybody. Yeah. No matter disabilities or anything like that. So making sure the colors are readable and things like that, like the blues.” (Agricultural economist at ERS)
- “It would be nice if once I've got into a data set I could kind of link out to the source there was where would I get the farm to school census? And then, I think this is probably me as a not very competent user, but I found someone in my state who used the farm to school census says, "Oh, what did they do?" And I could see who it was, but I really couldn't get their article. I'd have to go out and find that in a Agricola, or PubMed, or something, right? And then, I might not get the right article, so if somebody's already, on the backend, identified the articles, if there was a link to A DOI or something, it would be helpful to just identify that article.” (Full professor)
- “I think a thing that would be useful would be a cross search with two data sets, because most of these are probably can be used in ... RUCC codes are usually used in conjunction with other data. So those ones, in particular, I think would be something to think about.” (Agricultural economist at ERS)

3.5 Large Group Share: Summary of Reflection Question Responses

Following activity 2, all participants reconvened in a large group to discuss themes that emerged during the breakout sessions. The following questions were posed to the participants, and a summary of responses are included below:

1. What features of the dashboard did you find particularly useful or surprising?

- The information provided in each table has relevant information pertinent to who uses each dataset and where papers using those datasets are published.
- The use of filters allows for navigating the dashboard in different ways.
- Being able to track publications and authors, and then connecting publications is very helpful for researchers.
- The comprehensiveness of types of journals was interesting for researchers as it shows creative ways of using the data.

- The dashboard has value in that it can help individuals outside of the USDA learn about datasets that they might not otherwise be familiar with.

2. What limitations or problems did you encounter while navigating the dashboard?

- The interface seems very busy. A new user might have trouble finding things. A more simplified interface could make it easier to navigate the information
- Participants voiced that the list of authors were predominantly U.S.-based and wondered how other countries could be included.
- Linking authors to institutions and their emails would make it easier to contact them, as researchers often reach out directly to authors who use the datasets.

3. What features would you add to the dashboard?

- Participants indicated that they would want to know more about the datasets themselves. While finding publications and authors associated with the data is helpful, participants wanted to know more about the contents of the datasets.
- Several participants wanted to know how each dataset could be linked with other datasets. This comment speaks more to the functionality of the dashboard, as some raised concerns that it does not give information about how to use the dataset with other datasets.
- A clearer understanding of the subtopics of the publications could also help understand what is included in the main topic of the publication.
- Participants suggested that it would be helpful to know whether the data is restricted access or open access.
- If the dashboard point is to make datasets more accessible, comments from several participants suggested that providing metadata about the contents of the featured datasets would allow researchers to see inside the data rather than just who uses it and how it is published.

4. Would you use this tool in your own research?

- Most participants said they would use the tool but there were caveats about its current limitations and target audience.
- Participants indicated that the tool would be more useful for graduate students and early career researchers than for experienced researchers.
- Many participants viewed the dashboard primarily as a literature review tool in its current form.
- Participants suggested that adding more detailed information about datasets, including methods and potential for data merging, would increase its usefulness for research creativity.
- Several participants expressed a desire for more granular information about specific data items used within datasets, such as which questionnaire questions were most frequently used in research or the level of geographic scope at which the data was collected. This

level of detail was seen as important for understanding how datasets are applied in practice.

- Several participants expressed concern about the comprehensiveness of the tool, indicating they would not rely on it as an exhaustive source.
- Participants highlighted the need for better filtering of publications, specifically to distinguish between those that merely mention the data versus those that actually use it.
- Some participants, particularly those associated with ERS, noted the dashboard's importance for tracking data usage.
- Participants suggested that expanding the search capabilities beyond Scopus and including works in progress would improve the dashboard's utility.
- Several participants indicated that adding features like toggles for different document sources and user-contributed content would enhance the tool's usefulness.

4. Dashboard Highlights

4.1 Gaps Filled by the Dashboard

The workshop activities and discussions revealed several areas where the dashboard currently fills important gaps in the data usage landscape:

1. **Centralized Information Hub:** Participants consistently expressed a need for a single, comprehensive source of information about USDA datasets. In Group 2, a participant noted “Standard documentation across datasets (similar to how IPUMS provides documentation for multiple datasets)” as a desired feature, highlighting the need for a centralized platform. The dashboard addresses this by providing a centralized platform where users can find details about multiple datasets, their usage, and related research.
2. **Expert Identification:** The poll results showed that 77% of respondents found the feature identifying dataset-specific experts to be very or somewhat valuable. When participants were asked where they look to learn about datasets, a consistent response in the focus group was finding someone that works with the data.
 - In Group 2, a participant emphasized the value of “Detailed codebooks from data owners,” highlighting the importance of connecting with those who intimately understand the data.
 - Group 3 discussions indicated that knowing who publishes with the data is a feature of the ideal tool, validating the desire for direct connections with dataset experts.
 - In Group 4, when discussing how to learn about new datasets, one participant mentioned that they “go straight to the source to get the information - the data dictionary is particularly helpful (if it is publicly available),” and clarified, “I sometimes find the easiest way to find the information is to go straight to the source. I know using the food environment atlas, which has over 280 indicators of the food environment from a lot of different resources that we try to want to update it and evolve it. So we do sometimes look at other data sources and sometimes it's easy just to go straight to the source that pulls it, that makes it.”

These insights indicate that the expert identification available through the dashboard has the potential to fill a significant gap in the ability to easily connect with knowledgeable individuals for specific datasets.

3. **Research Discovery:** With 89% of respondents rating direct links to research articles as very or somewhat valuable, the dashboard fills a gap in helping users discover relevant research using specific datasets.
4. **Data Documentation Accessibility:** Many participants highlighted the challenge of finding comprehensive and accessible documentation for datasets. In Group 1, a participant noted that a key piece of information would be “to access data documentation and a variable list with definitions,” a suggestion that was received favorably by other participants in the group. In this sense, the existing dashboard can serve as a gateway to well-organized, standardized documentation across different USDA datasets.
5. **Usage Trends Visualization:** The dashboard’s ability to visualize dataset usage trends across different fields and over time fills a gap in understanding the broader impact and applications of USDA data. In Group 3, a participant suggested including “concentration of topical areas that use the data, such as 45% food insecurity, 20% health outcomes, etc...”, indicating interest in visualizing usage trends across fields. Given the existing metadata collected via the dashboard, these metrics can be integrated into future iterations of the dashboard.

4.2 Value to Different User Communities

While the workshop focused on research economists, the discussions highlighted the potential value of the dashboard to various user communities:

1. **Early Career Researchers:** The dashboard can be particularly valuable for graduate students and early career researchers, providing them with a quick way to understand the landscape of research using specific datasets and identify potential mentors or collaborators.
2. **Established Researchers:** For experienced researchers, the dashboard offers efficient ways to stay updated on new applications of familiar datasets and discover potential cross-disciplinary opportunities.
3. **Interdisciplinary Researchers:** The dashboard's ability to show dataset usage across different fields can facilitate interdisciplinary research by highlighting unexpected applications or combinations of datasets.

Although not directly represented in the workshop, the following groups may also find the dashboard valuable:

4. **USDA and Other Government Agencies:** The dashboard provides a tool for tracking the impact and reach of their data products, potentially informing decisions about future data collection efforts and resource allocation.

5. Policymakers: The dashboard could provide policymakers with insights into how USDA data is being used to inform policy decisions and identify areas where additional research might be needed.
6. Data Contributors: For stakeholders who contribute to USDA datasets (e.g., survey respondents), the dashboard could demonstrate the value and impact of their participation, potentially increasing engagement in future data collection efforts.

5. Suggestions for Improvement

Based on the feedback gathered from all four breakout groups, we have identified several potential improvements for the dashboard. These suggestions have been categorized into short-term improvements that require minimal to low investment ([Section 5.1](#)), and long-term developments that may require larger investments or funding ([Section 5.2](#)). The recommendations below represent the most frequently mentioned and highly valued features across all participant groups, addressing common pain points and desired enhancements. While some of these improvements may be more readily implementable, others require additional development, yet all are aimed at enhancing the utility and user experience of the dashboard.

5.1 Short-term Improvements (Minimal to Low Investment)

1. Describe Accessibility Information: Some datasets are not open access or require special permission for restricted use data. Providing these details to users would be a helpful enhancement. For each dataset, clearly indicate whether it's publicly available or restricted-access, and provide straightforward instructions on how to access the data.
2. Enhance Search Functionality: Participants indicated the importance of search functionality, including the ability to filter publications that actually use the data (versus those that merely mention it), and the option to search (using key words) within the dashboard itself. One recommendation is to implement keyword search capabilities within publication titles, abstracts, and dataset descriptions to help users find relevant research more efficiently.
3. Improve User Interface: Some participants remarked on the busyness of the interface and the information presented. One recommendation stemming from these comments is to enhance the dashboard's layout for better readability and navigation, including a more prominent placement of the dataset selection feature and a clearer way to reset filters which currently appear in the bottom left corner of the dashboard.
4. Improve Data Documentation Access: Participants consistently emphasized the need for clear, detailed, and easily accessible documentation for datasets, including information on data granularity, coverage, and access procedures. Our recommendation is to provide direct links to data dictionaries, codebooks, and user guides for each dataset featured on the dashboard. This functionality could lead to more granular filtering options, such as by geographic level (e.g., national, state, county) and time period covered by the dataset.

5.2 Long-term Developments (Larger Investment/Funding Required)

1. **Expand Dataset Coverage and Publication Sources:** The ability to easily find and access publications that have used a particular dataset is highly valued, as is identifying and contacting dataset-specific experts. Because not all data experts publish in the set of journals currently featured on the platform, one recommendation is to increase the number of USDA datasets featured on the dashboard by broadening the sources of publication and prioritizing those most frequently used by researchers and policymakers. The sources of publications can include other databases like Web of Science, Google Scholar, and agency-specific publication series.
2. **Provide Data Linkage Information:** There is a strong interest in understanding how datasets can be linked or merged with other data sources, as well as more information on the dataset's functionality and potential uses. The recommendation is to develop a feature that shows how different datasets have been combined or used together in previous research publications. This feature would: a) Highlight examples of successful data merges from published studies b) Provide information on common identifiers used for linking c) Discuss potential challenges researchers faced in combining datasets d) Offer insights into the methodologies used for successful data integration.
3. **Provide Usage Examples:** Participants suggested several user-friendly features, such as data visualizations, sample code, toggles for different document sources, and the ability for users to contribute their own work to the dashboard. One recommendation is to include sample code snippets or links to repositories (e.g., GitHub) with examples of how to load and clean data for common uses across different statistical software packages (e.g., R, Stata, SAS).
4. **Provide Personalized Recommendations:** As AI-based systems continue to evolve, search engine capabilities will continue to become more sophisticated. One recommendation that may be of interest is to develop an AI-driven system that can suggest relevant datasets or research based on a user's interests or research expertise.

6. Expanding User Engagement

The workshop also provided insights into how the dashboard can better serve a wider range of stakeholders and how its utility can be expanded through continued enhancements. This section outlines strategies for engaging diverse user communities and improving the dashboard based on user feedback.

6.1 Reaching Out to Different User Communities

The workshop highlighted the importance of engaging with various user groups to ensure the dashboard meets diverse needs. While the current focus has been on research economists, the insights gathered suggest potential value for other communities. Here are a few proposed events to reach these groups:

1. Graduate Student Workshops: Organize hands-on sessions introducing the dashboard in graduate-level research methods courses.
2. Data Contributor Appreciation Events: Organize events for data contributors (e.g., farmers, survey respondents) to show how their input is being used in research and policy, using the dashboard to illustrate impact.
3. Extension and Outreach Training: Host a training session for extension professionals on using the dashboard to access information relevant to their work with communities.
4. Policy Brief Series: Develop a webinar series, with each session focusing on individual topics using the dashboard to demonstrate how it can inform policymaking. This webinar could be targeted at government agencies whose datasets are featured on the dashboard.
5. Minority-Serving Institution Outreach: Conduct targeted outreach and training sessions at Historically Black Colleges and Universities (HBCUs), Hispanic-Serving Institutions (HSIs), and other minority-serving institutions to promote diverse engagement with the dashboard.
6. Black Farmers Association Workshop: Organize a specialized workshop for the Black Farmers Association to demonstrate how the dashboard can be used to access relevant agricultural data and research, potentially informing their advocacy and decision-making processes.

6.2 Defining Valuable Usage Metrics for Different User Communities

The workshop discussions highlighted the need for more comprehensive tracking of USDA dataset usage across various research fields and applications. Different user groups may have varying definitions of what constitutes valuable usage of a dataset. As one participant pointed out, “I would really try to concentrate on what you think the value added is to all of this.” This suggests a need for clear definitions of value that align with the goals of different user communities.

To cater to the needs of other target groups, the following metrics were suggested:

1. Research impact metrics: As noted in Group 4, “It would be helpful if we could filter within a filter. So, if we’re looking at say, Food Access Research Atlas and then filter within that, what other data sources were used in conjunction with that,” as often times datasets are used in tandem.
2. Policy impact metrics: Track citations in policy documents, legislative discussions, and real-world applications of research using the datasets. This would require expanding the publication sources beyond scholarly retrieval platforms.
3. Stakeholder engagement metrics: Develop ways to show how data contributions (e.g., from farmers participating in surveys) are being used in research and policy.
4. User diversity metrics: Track geographic and institutional diversity of dataset users to identify underserved communities or regions.

5. Dataset relevance metrics: Measure time from data release to first use to indicate the immediacy of a dataset's relevance.
6. Customizable metric dashboards: Allow users to define and track metrics most relevant to their specific needs or interests.

As ERS Administrator Spiro Stefanou noted, “Ultimately, we endeavor to measure the value of public data assets and the potential value of free public access to these data. That is a challenge that’s still on the horizon that we’re working towards.”

6.2 Potential Future Collaborations

The workshop discussions highlight the importance of collaborations to continue the development of the dashboard and optimize its utility and impact. Two avenues for partnerships have been identified, each offering distinct opportunities to enhance the tool’s functionality, scope, and relevance within the data-driven research and policy ecosystem.

Inter-agency Cooperative Frameworks:

Expanding collaborative efforts across federal agencies could significantly broaden the dashboard’s dataset coverage and applicability. By working across agencies, the development team can work on building standardized protocols for data documentation and usage metrics which may address shared challenges in data usage tracking and promotion. Inter-agency collaborations can also address the need for integrating data documentation requested by many of the participants.

Artificial Intelligence and Machine Learning Integration:

Exploring partnerships with initiatives like the National Artificial Intelligence Research Resource (NAIRR) Pilot presents opportunities to build AI technology underlying the collection of the metadata feeding into the dashboard. Additional applications of AI include developing intelligent data discovery algorithms and creating AI-driven recommendation systems for dataset combinations.

7. Conclusion and Next Steps

The workshop on the DemocratizingData.ai dashboard provided valuable insights into how the dashboard meets the needs and preferences of research economists who heavily use USDA datasets. These insights offer a foundation for refining and expanding the dashboard to better serve not only this core user group but also a broader range of stakeholders.

Key findings from the workshop include:

1. The high value placed on features that identify dataset-specific experts and direct links to research articles.
2. The need for more comprehensive and accessible documentation for datasets.

3. The importance of understanding dataset linkability and cross-dataset usage.
4. The potential for the dashboard to serve as a central hub for discovering and understanding USDA data usage across fields and sub-disciplines.

While the current iteration of the dashboard demonstrates that it has filled a gap in services offered to research economists, there are clear opportunities for enhancement. Short-term improvements, such as refining search functionality and improving data documentation access, can quickly increase the tool's utility for this target audience. Longer-term developments, including expanded dataset coverage and the presentation of data linkage information, will add considerable value to the dashboard for empirical researchers.

The workshop also highlighted the importance of expanding engagement beyond the current user base. By reaching out to diverse user communities - from early career researchers to policymakers and data contributors - the dashboard can evolve to meet a wider range of needs and maximize its impact.

Moving forward, the following steps are recommended:

1. Implement short-term improvements identified in the workshop.
2. Develop a roadmap for long-term improvements, considering both technical requirements and potential collaborations.
3. Design and execute a strategy for engaging with broader user communities, including hosting the events proposed in this report.
4. Establish a mechanism for collecting ongoing user feedback to ensure continuous improvement of the dashboard.
5. Look for collaborations to expand the dashboard's scope and impact.

The DemocratizingData.ai dashboard aligns with recommendations from the American Statistical Association (ASA) in their 2023 report "The Nation's Data at Risk: Meeting Americans' Information Needs for the 21st Century"[1]. Specifically, Recommendation 9 of the report (pg. 10 in the Executive Summary) urges parent agencies to interact with and support their statistical agencies, emphasizing the need for regular meetings between agency leadership to understand the statistical agency's functions, needs, and unique responsibilities. The dashboard addresses these needs by providing a centralized platform for discovering and understanding USDA data usage. It supports better engagement with the research community and improved access to federal statistical data, facilitating informed decision-making across government, business, and personal spheres.

References

[1] American Statistical Association. (2023). The Nation's Data at Risk: Meeting America's Information Needs for the 21st Century. A Report of the American Statistical Association in Partnership with George Mason University and Supported by the Sloan Foundation.
<https://www.amstat.org/policy-and-advocacy/the-nation's-data-at-risk-meeting-american's-information-needs-for-the-21st-century>

Appendices

A1. Method for Report Writing

The following steps outline our methods for generating the findings and insights presented in this report:

1. All four breakout sessions were recorded and subsequently transcribed verbatim using Rev.com. This process captured the entire content of the discussions.
2. The transcripts and other workshop materials were then analyzed to summarize the key outcomes of each activity:
 - Activity 1: The ideas generated by participants using the IdeaBoardz online collaboration tool were collated and categorized. This included analyzing the pieces of information participants deemed most important, the tools and resources they currently use or would like to have, and the features they desire in an ideal data usage tracking tool.
 - Quick Poll Results: The responses to the quick poll were summarized, providing quantitative data on participants' perceived value of specific features of the dashboard.
 - Activity 2: The impressions and feedback from each group following the dashboard demonstration were summarized. This included noting common themes, insights, and specific suggestions for improvement across all four breakout groups.
3. The information from all activities was then synthesized to identify overarching themes, common pain points, and shared desires among the participants. This synthesis forms the basis of the key findings and insights presented in this report.
4. The initial findings were reviewed by Nick Pallotta and Julia Lane. This report includes their feedback.

A2. Workshop Details

Date: Monday, June 10, 2024

Time: 12:00 PM - 1:30 PM Eastern

Location: Zoom

Website: <https://usda-data-usage-workshop.netlify.app/>

Agenda

12:00 - 12:10 PM: Welcome Remarks

12:10 - 1:05 PM: Dashboard Demo and Activities

- Presentation of the DemocratizingData.ai dashboard
- Interactive activities in breakout rooms

1:05 - 1:25 PM: Group Share

- Reconvene in main room
- Discussion from breakout sessions

1:25 - 1:30 PM: Concluding Remarks

A3. IdeaBoardz

Excluded for confidentiality purposes.

A4. Breakout Session Transcripts

Excluded for confidentiality purposes.

A5. Participant List

Excluded for confidentiality purposes.