NORC at the University of Chicago

EXECUTIVE SUMMARY

September 2025

DUP-23-N04: Data Usage Platform as a Federal Data Asset

Presented by:
NORC at the University of Chicago

Presented to:
NCSES
May Aydin

The America's DataHub Consortium (ADC), a public-private partnership is being utilized to implement research opportunities that support the strategic objectives of the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF). This report documents research funded through the ADC and is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any opinions, findings, conclusions, or recommendations expressed in this report do not necessarily reflect the views of NCSES, NSF, or their government partners. Please send questions to ncsesweb@nsf.gov. This product has been reviewed for unauthorized disclosure of confidential information under NCSES-DRN25-063.

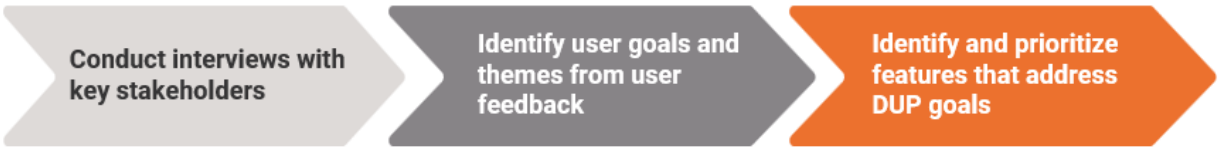
Project Background

NORC was contracted by NCSES to conduct user research and develop a data usage platform prototype which could be adopted by a potential, future NSDS. The prototype in this project was designed to assist a broad set of stakeholders in understanding how federal data assets are used for research, reporting, planning, and decision-making. Understanding the uses of federal data can support collaboration across various research and policy communities and promote cross-disciplinary approaches to research and evidence-based policy making. Highlighting uses and applications of federal data assets can also support federal agencies' business needs and decision-making related to stakeholder requirements and stakeholder engagement while supporting investments in data programs.

Methodology

NORC and our partner, Fuzzy Math, (the "NORC team") analyzed existing data sources and dashboards and structured user interviews to discover user needs, which then informed the features of the DUP prototype to ensure that the tool addresses user needs and gaps in the data ecosystem and strategically leverages available data sources.

Figure 1: Methodology Overview



Environmental Scan

The NORC team conducted a review of relevant dashboards focused on visuals and the user experience, including intuitive navigation, clarity in data presentation, and flexibility in interface design. In addition to evaluating other federal entities' dashboards, we identified and reviewed several commercial applications developed outside of the federal government for visual design and analytic best practices. Finally, during user interviews, participants were shown examples of current data usage dashboards to obtain their feedback on the overall usability and value of the dashboards.

Key Findings – UX and Visual Design Review

This section summarizes key findings from both user interviews and an extensive UX and visual design review of existing data dashboards conducted by the NORC team. The goal of this review was to

identify best practices that should be employed in the development of a DUP to increase engagement and positive user experiences with the platform.

User Experience

- **Filtering and sorting:** A DUP should prioritize robust filtering and sorting capabilities to empower users in exploring and analyzing data according to their specific needs.
- **Export options:** Offering diverse export options, including PDFs, images, PowerPoint, and data formats, ensures flexibility for users to view and share information.
- **Decrease cognitive load:** Providing a concise summary at the top of the page offers users an immediate overview, aiding quick comprehension.
- **Undo and redo functionality:** Incorporating undo and redo functions enhances user control and flexibility during data exploration.
- **Accessible language for non-technical users:** Non-technical, succinct language supports comprehension and aligns with best practices in user-centric design.
- **Exploratory features:** Platforms should incorporate features that facilitate exploratory data analysis, allowing users to delve deeper into datasets.

Visual Design

- **Icons used for graphic interest and function:** The distinction between icons used for graphic interest and those with a functional purpose should be clear.
- **Alerts styled with intuitive color palettes:** Alerts should be styled with intuitive color coding, such as green for a success confirmation or blue for a general announcement.

Accessibility

- **Visual harmony for a user-friendly presentation:** Simplifying the layout and organizing information effectively, including a clear hierarchy, distinct headers, and clear labels, contributes to a user-friendly experience.
- **Variety of typographic scale and weight:** Using a variety of typographic scales and weights, including bold, medium, and regular/book text weights, can help enhance overall readability.
- **Small font sizes:** Font sizes of 12px or smaller are not accessible to users. Following Web Content Accessibility Guidelines (WCAG) best practices, 16px is recommended for body copy text.
- **Branding distinctiveness:** The DUP should enhance branding by introducing a variety of distinctive elements and features, fostering a strong visual identity that differentiates the platform.

User Experience

- **Guidance and alerts:** Providing guidance through tooltips, documentation, and clear error messages enhances the overall user experience.
- **Customization options:** Customizable dashboards are essential for tailoring the user experience to individual preferences and needs.
- **Interactivity and responsiveness:** Enabling interactive elements that respond promptly to user interactions is fundamental.

Key Findings – Stakeholder and User Interviews

NORC conducted interviews with a variety of federal employees to understand the type of data they collect, how they measure both internal and external data usage, and what they would like to know about how their data are used. NORC interviewed both statistical and non-statistical agencies. NORC also completed interviews with non-federal data users to understand how they use federal data, share findings derived from their analyses, engage with data platforms, and their perceptions of the value of tracking data usage. Participants included a graduate student, data journalists, staff supporting policymaking, and researchers. Table 1 summarizes the goals for each group interviewed as well as suggested features for inclusion in a DUP.

Table 1: Examples of needs by persona and corresponding features proposed for a DUP

Persona	Need	Feature
Federal (statistical agency)	How do data-related requests compare to usage statistics?	Time explicit usage statistics at the data asset level
Federal (non-statistical agency)	Which data assets are used alongside data from other agencies?	Integrated usage statistic tracking across agencies
Non-federal (researcher)	What topics of study do data products support?	Publication data with topic tags
Non-federal (policymaker)	How has an update to a data asset impacted its usage?	Time explicit usage statistics (for pre- and post-intervention comparison)
Non-federal (data journalist)	What data assets are the best sources of information on specific topics?	Most frequently used federal data sources by topic

Sources of Usage data

NORC completed an environmental scan of potential sources for usage data to aggregate data usage metrics that reflect the priorities and research questions identified in stakeholder interviews. This scan of potential data sources considered the volume of available data, type of information provided, licensing options, cost, and relevance to general agency-level data tracking needs. Based on factors like coverage, accessibility, cost, and scalability, NORC decided to collect usage data from the following sources in our DUP prototype:

- Citation usage data from OpenAlex
- Policy document usage data from Overton
- Restricted access applications from the Standard Application Process (SAP)
- Counts of data views and downloads from Data.gov

Policy Usage data

NORC assessed the cost, effort, and coverage considerations for developing a custom software solution to collect policy usage data compared to sourcing policy usage data from Overton. Overton collects and curates roughly 12 million policy documents from 1,700+ sources and is automatically updated as new documents are released from these sources. Data are available through an online interface, data dumps, and an application programming interface (API). Based on ease of access, low start-up costs, and coverage of data, NORC sourced policy usage data for its DUP prototype from Overton.

Media Usage data

While media usage presents is a potentially valuable dimension of data usage, the NORC team decided not to include media usage data in its DUP prototype due to two key factors:

- None of the sources we evaluated appeared to offer free or low-cost programmatic search options offering coverage across multiple news sources.
- Only one of the sources we reviewed provided readership analytics, such as the number of times an article was viewed. This information is instrumental in identifying publications and policy documents that cite data assets as well as their citation information (e.g., how many times the citing publication has been cited by others) and need parallel information for news articles to help draw comparisons of engagement and reach (e.g., how many times the citing news article has been read) across the sources included in the platform.

For a future DUP, NORC recommends continued exploration of the viability of including media usage data, especially if an API becomes available that provides reader analytic data beyond article text.

DUP Prototype

The NORC prototype is a free-standing R application, developed by NORC methodologists and informed by a visual design framework developed by Fuzzy Math based on extensive user interviews completed in 2024. The prototype includes real usage data for the following three agencies: the Bureau of Justice Statistics (BJS), the U.S. Census Bureau, and the Internal Revenue Service Statistics of Income (SOI) program.

The prototype includes data representing multiple stages of the data access and usage continuum: from data exploration and acquisition to use in academic publications and policy documents. Within the DUP prototype that NORC built, users can explore:

- How often data products are viewed or downloaded from Data.gov
- How many applications have been submitted or approved for restricted data
- How often data products are cited in scholarly publications or policy documents

In addition, users of NORC's DUP prototype can explore the most cited data products by agency, compare usage metrics between agencies, track data downloads by agency over time, and explore the title, themes, and sources of data usage in both academic publications and policy documents.

In addition to a data exploration component, enabled by an interactive dashboard of data visualizations and tables that allow for sorting and filtering, the NORC DUP prototype also includes a detailed glossary that defines key words and metrics and contextualizes them within the purpose of the DUP. An FAQ page further elucidates common questions that DUP users have, such as:

- How often is the data updated within the DUP prototype?
- Which metrics are available for each data product?
- Why are some metrics missing for certain data products?

Future Recommendations

In support of a scalable, centralized DUP that can function as a shared service, NORC recommends the following enhancements for a future, production-ready DUP:

- **Containerization**. Containerization is an approach to packaging software such that it has the ability to run in a wide variety of computing environments, including on different operating systems. In essence, containerization would "bundle" the DUP, its required software (e.g., R or Python), and any relevant dependencies to ensure the DUP is able to run. A containerized DUP could then run across operating systems and computing infrastructures.
- **Cloud-based infrastructure.** Hosting the DUP in a cloud environment offers the flexibility needed for the DUP to function as a shared service. A cloud-hosted DUP would enable agencies to offload administrative overhead when possible while retaining the ability to launch private instances when

necessary (e.g., for security and/or legal reasons). This cloud-hosted approach would enable agencies to benefit from the economy of scale, in terms of administration and costs, while retaining strict control over security and data access—which is necessary for tiered access.

- **Automating data ingestion.** NORC's DUP prototype relies upon application programming interfaces (APIs) and webscraping for collecting data from most sources. As such, this code could be automated to collect and process data at regular intervals, which would enable the DUP to present real-time or near real-time usage data. Software, such as Rundeck, Jenkins, or even cloud-native solutions depending upon the cloud platform, would allow this data collection and ingestion process to run at regular intervals.