



# Interagency Council on Statistical Policy

Leaders of the United States Federal Statistical System

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

Prepared for the Interagency Council on Statistical Policy  
By the ICSP Private Data Steering Group

Summary Report

January 6, 2023



# Table of Contents

<b>Introduction</b> .....	<b>3</b>
<b>Approach</b> .....	<b>4</b>
<b>Private Data Case Studies</b> .....	<b>6</b>
<b>Report Findings</b> .....	<b>7</b>
Reliance on private datasets.....	8
Reasons for use of private datasets.....	9
Private data acquisition—challenges and practices .....	14
Private data preparation and analysis—challenges and practices .....	19
Sustainability of private data use—risks.....	25
Best practices and lessons learned in the acquisition, preparation, analysis, and sustainability of private data .....	27
Spinoff benefits.....	30
<b>Conclusions</b> .....	<b>32</b>
<b>Appendix: Survey Instruments</b> .....	<b>34</b>
General Survey.....	34
Case Survey .....	41



## Introduction

The Interagency Council on Statistical Policy (ICSP) advises and assists the Chief Statistician of the United States<sup>1</sup> in the coordination of the Federal statistical system; the implementation of statistical policies, principles, standards, and guidelines; and the assessment of statistical program performance. Chaired by the Chief Statistician of the United States, the ICSP is comprised of 26 members. Pursuant to the Paperwork Reduction Act of 1995 (PRA), all 13 heads of the principal statistical agencies and units are members. Pursuant to the Foundations for Evidence-based Policymaking Act of 2018, all 24 Statistical Officials are also members; however, 11 of the Statistical Officials are also heads of principal statistical agencies or units.

For nearly a century, federal information collection to inform policymaking has been based largely on household and establishment surveys. In recent years, however, three factors have caused ICSP members to reevaluate their reliance on survey-based data collections: (1) falling survey response rates; (2) advances in information technology that enable far greater access to administrative records, particularly from private (nongovernmental) sources; and (3) the requirements of the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) to specify and then collect new information needed to intelligently guide public policy.

To date, ICSP members have engaged in uses of private data independently of other agencies and the ICSP as a whole. Considering the necessities and opportunities for using private data in the future, including the interest in high-frequency data, the ICSP created a small working group to gather information from ICSP members regarding their experiences.<sup>2</sup> Specifically, the goal of this research project was to ascertain the following:

- the breadth and nature of statistical agency uses of private sector datasets,
- the benefits and challenges in the use of these datasets, and
- the implications of the findings for federal uses of such datasets going forward.

---

<sup>1</sup> The Chief Statistician of the United States is appointed by the Director of OMB to carry out the Director's statistical policy and coordination activities under the Paperwork Reduction Act of 1995.

<sup>2</sup> The ICSP Private Data Steering Group is chaired by Erich Strassner, formerly Associate Director for National Economic Accounts at the Bureau of Economic Analysis (BEA) (through November 4, 2022). Members include William Wiatrowski, Deputy Commissioner, Bureau of Labor Statistics; John Staub, Director, Office of Energy Production, Conversion, and Delivery, Energy Information Administration; and Annabel Jouard, Economist, BEA (through March 14, 2022). Andrew Reamer, Research Professor at the George Washington Institute of Public Policy, George Washington (GW) University, staffed the steering group under a grant from the Alfred P. Sloan Foundation, with research assistance from GW Ph.D. candidate Danielle Gilmore (through November 15, 2021).



## Approach

The project collected information from ICSP members through two sequential surveys. The first (“general survey”), sent to each ICSP member agency in July 2021, asked a series of questions regarding agency utilization of and experience with **private sector data**<sup>3</sup>, defined as “data purchased or licensed from private-sector entities within the past five years.” Question topics included the extent of reliance on private data; reasons for use; challenges in acquisition, analysis, and sustainability of use; responses to challenges (solutions and best practices); and approaches to dataset management. The survey also asked each agency to “describe up to three cases of your agency’s experiences with current or planned uses of private data.” Responses were collected through the Qualtrics survey platform. (The general survey instrument is provided in the [Appendix](#).)

The second survey (“case survey”) asked 8 ICSP agencies to provide detailed information on 12 case examples described in the general survey (The case survey instrument also is provided in the [Appendix](#).)

Twenty ICSP member agencies responded to the general survey. Fourteen respondents identified themselves (see the box “Identified ICSP Respondents to General Survey”). Of the 14 identified respondents, 12 are principal federal statistical agencies. (The two exceptions are the Department of Veterans Affairs and the U.S. Agency for International Development.)

---

<sup>3</sup> Those providing responses for this project identified some information that didn’t precisely meet the definition. All reported information was included in this report.



### **Identified ICSP Respondents to General Survey**

- Bureau of Economic Analysis (BEA), Department of Commerce
- Bureau of Justice Statistics (BJS), Department of Justice
- Bureau of Labor Statistics (BLS), Department of Labor
- Bureau of Transportation Statistics (BTS), Department of Transportation
- Census Bureau (Census), Department of Commerce
- Economic Research Service (ERS), Department of Agriculture
- Energy Information Administration (EIA), Department of Energy
- National Agricultural Statistics Service (NASS), Department of Agriculture
- National Center for Education Statistics (NCES), Department of Education
- National Center for Health Statistics (NCHS), Department of Health and Human Services
- National Center for Science and Engineering Statistics (NCSES), National Science Foundation
- Office of Data Governance and Analytics, Department of Veterans Affairs (VA)
- Research, Statistics, and Policy Analysis, Social Security Administration
- U.S. Agency for International Development (USAID)

In August 2021, the case survey was sent to eight agencies (BEA, BLS, BTS, Census, NASS, NCES, NCHS, and the VA) seeking information on 12 cases selected from over 30 nominated in the general survey. The cases were chosen to represent diverse sets of agencies, data providers, and uses. Eleven completed case study surveys were received from seven agencies. The basic characteristics of each case study are summarized in the table [“Private Data Case Studies.”](#)

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



## Private Data Case Studies

Agency	Federal data program	Official (O) or exploratory (E) private data use	Nature of dataset	Individual (I) records or aggregated (A)	Use
BEA	Gross domestic product and other national accounts (quarterly)	O	Payment card transactions	I	Inform a ML-based prediction tool for Quarterly Services Survey
BLS	Consumer Price Index (CPI)	O	New motor vehicle transactions	I	New car prices in CPI
BLS-Census-BEA	Exploration of possible uses	E	Payment card transactions and payroll processing	A	Exploration of potential statistical agency uses
BTS	Daily transportation indicators (COVID-19)	E	Cell phone location data	A	Track volumes of daily travel during COVID-19 pandemic
Census	Administrative Records Census Experiment	E	Personal demographic characteristics	I	To assess value in supplementing current data sources (particularly covering persons otherwise missed and providing more timely data) that together could serve as sampling frame and replace some household data collection
Census	American Community Survey (ACS)	O	Household status and characteristics	I	In combination with government data, help identify vacant housing units.
NASS	Cropland Data Layer, disaster analysis	O, E	Satellite monitoring—crop acreage, disaster assessment	I	In conjunction with NASS data products, assess impacts of extreme weather events and anomalies
NCHS	National Hospital Care Survey (NHCS)	O	Data elements similar to those collected by NHCS	I	Augment low-response NHCS data collection to generate statistically sound nationally representative estimates
VA	Veteran Population	O	Veterans' demographic, socioeconomic, and geographic characteristics	I	Include veterans who do not use VA services or benefits
VA	Veteran Population	O	Prescription drug monitoring	I	Monitoring opioid use among veterans.
VA	Veteran Population	O	Economic projections	I	Enable projections of veteran population at lower levels of geography



## Report Findings

The results of the two surveys are described in detail in the remainder of this report in the following sections:

- Reliance on private datasets
- Reasons for use of private datasets
- Private data acquisition—challenges and practices
- Private data preparation and analysis—challenges and practices
- Sustainability of private data use—risks
- Best practices and lessons learned in the acquisition, preparation, analysis, and sustainability of private data
- Spinoff benefits

### Reliance on private datasets

Almost all responding ICSP agencies have used private data in the last 5 years. The following table provides a sample of some of the responses.

Agency	Number of private datasets in use	Percentage of statistical programs using private data
BEA	142	100
EIA	≈ 80	100
USAID	50	> 95
Census	≈ 20	20
NCES	12	30
BTS	10	50
NCHS	≈ 8	40–50
BLS	≈ 5	10

- Three agencies—BEA, EIA, and USAID—indicate **extensive reliance** on private datasets in both nominal and relative terms. Both the number of private datasets these agencies use (142, ≈ 80, and 50, respectively) and the percentage of their statistical

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



programs using private data (100, 100, and > 95, respectively) are far greater than the number reported by other agencies. Across the 12 respondents, these 3 agencies account for over 80 percent of the private datasets used.

- Four agencies—Census, NCES, BTS, and NCHS—report **moderate reliance** compared to the other two groups—between 8 and 20 private datasets and between 20 and 50 percent of statistical programs.

**The extensive reliance on private data by each of the three agencies in the first group is of long standing and rooted in each agency’s unique history.**

- Under Commerce Secretary Herbert Hoover in the Harding-Coolidge Administrations, BEA’s predecessor—the Statistical Research Division of the Bureau of Foreign and Domestic Commerce (BFDC)—created a “Clearing House of Business Statistics,” built primarily on private datasets from trade associations, to generate the monthly numbers published in the *Survey of Current Business*.<sup>4</sup> BFDC’s inventory of private data enabled its initial determinations of national income (in 1933) and gross national product (in 1942).<sup>5</sup>
- In recognition of the vulnerability of the U.S. economy to energy shocks of various types, EIA and its predecessor were created in the 1970s to generate detailed, current data on multiple types of energy inputs and outputs, as well as forecasts of various kinds, to guide federal energy policy and regulation.<sup>6</sup> At that time, the U.S. private energy sector had a substantial institutional infrastructure, in the form of trade associations and related research entities (for example, the Electric Power Research Institute), that has enabled EIA to gather substantial amounts of data from private sources. In addition, data on energy inputs and outputs lend themselves to

---

<sup>4</sup> See “[Clearing House of Business Statistics](#),” *Survey of Current Business* 11 (July 1931).

<sup>5</sup> See Andrew D. Reamer, “[The Origins of the Survey of Current Business: A Window on the Evolution of Economic Policy, Research, and Statistics](#),” *Survey* 100 (October 2020). Hoover modeled the BFDC Clearing House on the Statistical Clearing House created during World War I by the Central Bureau of Planning and Statistics to support the efforts of the War Industries Board, the War Trade Board, and the U.S. Shipping Board.

<sup>6</sup> Congressional Research Service (CRS), [The U.S. Energy Information Administration](#), CRS report R46524 (Washington, DC: CRS, September 9, 2020).



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



being comprehensively gathered by private entities than do inputs and outputs in most other sectors.<sup>7</sup>

- As USAID’s datasets are international in scope, the agency makes clear that it must rely on secondary data sources.<sup>8</sup>

## Reasons for use of private datasets

The general survey and the case survey each asked the ICSP agency to indicate its reasons for using private data. The multiple choices offered were the same in both surveys, with slight variation in wording. The case survey respondents also were given the opportunity to add text explanations. For each survey, the choices offered and the distribution of responses, rank ordered by frequency, are provided below.

General survey: Over the past five years, for what reasons has your agency used private data? (Check all that apply)	Number	Percent
To supplement or combine with existing agency held data	14	82.4
To better understand other indicators of the economic environment (situational awareness)	12	70.6
Verification, quality control or quality assurance for existing data or estimates	9	52.9
To continue current reporting capacity of agency priorities	9	52.9
For use as or to identify a survey frame	8	47.1
To make available to agency stakeholders and data users (publish data to the web)	8	47.1
Direct data collection is too costly (private data sets are a cost-efficient alternative)	7	41.2
Direct data collection is too burdensome on respondents (resulting in poor response or data quality)	4	23.5
Number of respondents	17	

<sup>7</sup> EIA and NASS are the only two principal federal statistical agencies with a focus on readily quantifiable outputs. In contrast to EIA, NASS has relied on survey collections since its founding in the 1860s.

<sup>8</sup> For information on USAID data sources, see [USAID IDEA \(International Data and Economic Analysis\) data sources](#), [“About” USAID IDEA](#), and [USAID Source and Indicator Guide](#).

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



Case survey: Describe how these private data are or will be used at your organization (Check all that apply)	Number	Percent
To supplement or combine with existing agency held data	7	63.6
To make available to agency stakeholders and data users (publish data to the web)	7	63.6
To better understand other indicators of the economic environment (situational awareness)	6	54.5
In lieu of direct data collection, which is too costly (private data are a cost-efficient alternative)	6	54.5
Verification, quality control or quality assurance for existing data or estimates	5	45.5
To continue current reporting capacity of agency priorities	4	36.4
In lieu of direct data collection, which is too burdensome on respondents (resulting in poor response or data quality)	4	36.4
Other (Specify)	3	27.3
For use as, or to identify, a survey frame	2	18.2
Number of respondents	11	

An examination of the responses yields the following several observations:

- Most ICSP agencies have multiple reasons for using private data.** In the general survey, 14 of 17 agencies indicated 3 or more reasons. The mode was three (seven agencies). In the case survey, agencies checked 3 or more reasons in 9 of the 11 cases.
- Eight reasons were chosen by a significant portion of private-data-using ICSP agencies.** In the general survey, each reason was selected by over a quarter of these agencies; four reasons were checked by over half.
- “Supplement or combine with existing agency-held data” was the single most frequently cited reason.** Additional reasons selected with some frequency included enhanced data offerings, enhanced data analysis, quality control, sampling frame, maintenance of agency capacities, reducing cost, and mitigating low survey response rates.
- ICSP agencies with extensive or moderate reliance on private data did not select significantly more reasons than did agencies with modest reliance.** On average, 5.0 reasons were selected by the seven agencies in the first group and 4.6 reasons by the five agencies in the second group.



As can be seen in the box “Reasons for Private Data Use in Case Studies,” the text addenda in the case survey offer greater specificity on the diverse reasons for using private data.

### **Additional Reasons for Private Data Use in Case Studies**

- Bridge timeliness gap of survey data.
- Access to actual transaction prices and a much greater number of observations than provided by direct data collection.
- Experiment to see if mobile device data could replace travel surveys.
- Broad and open-ended exploration of potential uses for private administrative data.
- Complement government administrative data sources.
- Level of data granularity not otherwise available.
- Augment low-response rate to generate statistically sound nationally representative estimates.
- Provide data on veterans who have not used VA services and benefits.
- Provide data on incarcerated veterans not available otherwise.
- Project veterans’ population at a lower level of geography.

## **Private data acquisition—challenges and practices**

The general survey asked each ICSP agency to indicate the types of challenges it experienced in acquiring private data. Multiple choices were offered, including an “Other” category.

<b>General survey: Over the past five years, what were any challenges in obtaining private data? (Check all that apply)</b>	<b>Number</b>	<b>Percent</b>
One or more challenges	16	84.2
Cost	13	81.3
Legal hurdles	9	56.3
Other (Specify)	6	37.5
Difficulty obtaining sufficiently granular data	6	37.5
Security challenges in accessing data once obtained	4	25.0
Technological challenges in accessing data once obtained	4	25.0
No challenges	3	15.8
Number of respondents	19	



An examination of the multiple-choice responses indicates the following:

- **The large majority of ICSP agencies that use private data identified one or more challenges in private data acquisition.**
- **Among the 16 agencies experiencing data acquisition challenges, 6 said they faced 4 or more challenges.**
- **Cost and legal hurdles were the most frequently cited challenges.**
- **Agencies also expressed concern about the same dataset being purchased by different units in the agency.**

In the text section, agencies identified specific challenges:

### **Observations on Acquisition Challenges**

- Backlog in federal contracting officer processing of competitive bid acquisitions.
- Inadequate methodology documentation and data transparency.
- Federal set-aside for small business contracts—can impede obtaining data from large firms.
- Vendors proprietary tools. These vendor tools are very limiting, come with licensing agreements, and cost for maintenance.
- Vendor security protocols. In many instances, these vendors want access inside the agency firewalls to make updates to their tools and infrastructure. They do not want the agency to have insight to the process.

In the case survey responses, additional detail on acquisition challenges was provided.



**Case Survey Respondents on Acquisition Challenges**

- These data are often expensive, which could present challenges.
- Contractual obligations to obtaining these data could change, causing operational risks (changes in vendors, scheduled deliveries, formats, etc.)
- Securing funds to purchase data is inconsistent from year to year.

The general survey asked each ICSP agency about the nature of its approach to managing the acquisition of private datasets.

- **The nature of approach varies across the 19 responding agencies** as follows:
  - Six indicated they take a centralized approach to managing private data contracts
  - Six indicated a decentralized approach
  - Three checked both approaches
  - Three checked neither and said their approach varies by project
- **Almost half of agencies (7 of 15) that checked a centralized or/and a decentralized approach indicated some flexibility**, also checking “Varies by project.”

<b>General survey: What are your agency’s approach(es) to managing private data contracts or memorandums of understanding (MOUs) for private data? (Check all that apply)</b>	<b>Number</b>	<b>Percent</b>
Varies by project	10	52.6
Decentralized (at sub-agency level, such as individual program)	9	47.4
Centrally (at the agency level)	9	47.4
Limited to a specified dollar amount	3	15.8
Other (Specify)	2	10.5
Number of respondents	19	

The general survey also asked agencies about planned and opportunistic acquisitions of private data.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- **A majority (11 of 19) have specified the use of private data in an agency management document** such as a congressional budget justification, strategic plan, data strategy, or operational plan.
- **Almost half (9 of 19) said that they obtained private data at least once in the last 5 years in response to a specific opportunity or need—outside of any plan, strategy, or budget initiative.** Examples given include acquisition of datasets in response to unforeseen emergencies (COVID–19 pandemic and energy related, for example), low survey response rates, new awareness of available private data, and a need for bibliometric data.

Ten case survey respondents provided details of various aspects of their data acquisition process. (In one case, the data had not yet been obtained).

- **Over half of the agencies conducted a quality assessment before acquiring the data.** Quality assessment was carried out in six cases and was not carried out in four.
- **There are correlations between the nature of the dataset (customized, or “off the shelf”), the means of obtaining it (legal agreement or credit card purchase), and its cost.**
  - In most cases, the agency had a legal agreement to purchase a customized dataset at a cost that ranged to over \$1 million. In several of these cases, the customized dataset was a specified extract of a larger standard dataset. And in three of those cases, the agency also purchased “off-the-shelf” data.
- **In every case, the purchase of private data had terms and conditions.** The distribution of the types of terms and conditions can be seen in the table below.
  - Terms and conditions included term licensing, vendor perpetual ownership, and a requirement to publicize the vendor's name.
  - Regarding terms and conditions in the “Other” category, the following are of note:
    - In one case, while there was no legal agreement as such, the agencies agreed to adhere to the specific conditions set by each of the primary data sources.
    - One agency could publicize the direct source but not the underlying sources.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- In one case, the agency was not allowed to publish quality metrics that compared private data providers.
- In one case, the agency could not include one private dataset in public use or restricted files.
- In one case, the agency agreed to delete certain historical data at the end of the agreement.

Case survey: What terms and conditions, if any, were imposed by the private data provider? (Check all that apply)	Number	Percent
Other (Specify)	8	72.7
Term licensing	5	45.5
Perpetual ownership	4	36.4
Requirement to publicize vendor name	3	27.3
Republication rights [verbatim, aggregated, only images of a chart, none]	0	0.0
Site-wide licenses to reduce copyright infringement within the organization	0	0.0
Inability to publicize vendor name	0	0.0
No terms and conditions	0	0.0
Number of respondents	11	

## Private data preparation and analysis—challenges and practices

The general survey asked each ICSP agency to indicate the types of challenges it experienced in analyzing or using the private data. Multiple choices were offered, including an “Other” category.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



General survey: Over the past five years, what were any challenges in analyzing or using private data once obtained? (Check all that apply)	Number	Percent
Experienced challenges	16	84.2
Methodology is not clearly documented (collection or aggregation methods are not clear)	14	93.3
Data require cleaning or additional manipulation to make useful	12	80.0
Data are of poor quality or incomplete	11	73.3
Data do not fully capture or match agency measurement objectives (construct validity)	11	73.3
Data are not fully representative of the measurement universe (agency frame)	10	66.7
Data are not easily matched to existing agency data	9	60.0
Other	3	20.0
No challenges	3	16.7
Number of respondents	19	

An examination of the multiple-choice responses indicates the following:

- **The large majority of ICSP agencies that use private data identified one or more challenges in private data analysis.** Only 3 of 19 respondents said they did not experience any data analysis challenges.
- **Most agencies said they faced a significant number of analytic challenges.** Among the 16 agencies experiencing data acquisition challenges, 11 said they faced 4 or more challenges. Again, the number of challenges faced does not strongly correlate with the extent of an agency’s reliance on private data.
- **Every type of challenge was frequently cited.** Each of the six types of data analytic challenges was cited by nine or more agencies.

The three agencies indicating an “Other” challenge wrote the following:

- Time period of data not clear; methodology changed over time.
- Lack of transparency hinders quality assessment and knowing how to use and link with other data and analysis.
- Data are lacking content for age 18 and under population.



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



For context, **nearly three-quarters of the respondents (14 of 19) said they have adopted a set of data principles or criteria regarding the use of private data.** There is no apparent correlation between the presence of principles or criteria and the absence of challenges.

With 1 being very easy and 5 being very difficult, case survey respondents gave 11 private datasets an average rating of 3.3 regarding the level of effort required “to get these data to a point where analysts could review or test these data.” Five datasets received a 3, two received a 4, and one received a 5. One dataset received a 1 and one received a 2.

Case survey respondents then indicated the methods used to assess data quality (table below). On average, each respondent used two methods. Comparability to existing data and availability of detail of provider’s methodology were each cited by most respondents.

Case survey: Once acquired, how did/does your organization assess data quality? (Check all that apply)	Number	Percent
Comparability to existing data	8	72.7
Availability of detail of provider's methodology	6	54.5
Formal quality assessment methods	4	36.4
Compatibility to traditional official statistics	2	18.2
Other (Specify)	2	18.2
Number of respondents	11	5.3

In 6 of 11 cases, the agency says it would be willing to share any written quality assessment process or model.

For the case studies, information was requested on the acquisition costs, months needed from acquisition to preparation and from preparation to analysis, and the annual license or maintenance fee. In general, it appears, the months required for data preparation are relatively minimal; the months required from preparation to analysis are more significant in those cases that reported a number; and both initial cost and annual maintenance costs varied greatly. The following tables provide a sense of the range of costs and time involved in the case studies.



**Number of cases by range of initial and ongoing costs**

Range of costs for case study data	Initial cost	Ongoing annual cost
\$0	1	5
\$1-\$250K	2	2
\$251K-\$500K	4	2
\$501K-\$1M	0	0
Greater than \$1M	1	0
Not provided	3	2

**Number of cases by range of time for initial preparation and analysis**

Range of time working with data	Initial preparation	Analysis
Less than 2 months	6	1
2-5 months	1	0
6-11 months	1	1
1 year or more	0	2
Not provided	3	7

The nine datasets that advanced to production received an average rating of 2.4 regarding the level of effort required “for staff or contractors to use these data each time.” Two datasets received a 1, three received a 2, three received a 3, and one received a 5.

In text remarks, respondents for eight cases noted specific challenges in data preparation and analysis (see the box “Case Survey Respondents on Preparation and Analytic Challenges”).

**Case Survey Respondents on Preparation and Analytic Challenges**

- There certainly are limitations with the data, including some concerns with coverage and representativeness. Also, users expressed concerns with contradictions with official estimates/differences from other stakeholder or agency data and concerns with accuracy relative to official alternatives.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- Agency recognizes that the “learning phase” that begins once private data is acquired and ends once it’s used in the official index estimates is an expensive time because we are paying for the private data and have expenses for direct collection at the same time. The cost of direct collection ends only when agency can rely on the private data for the calculation of published estimate. [A]nalysis for actually using this data to calculate experimental index estimates took approximately seven years.
- Cell phone location
  - Agency launched a minimally viable statistical product quickly without fully understanding the data to meet demands of senior officials for statistics “yesterday”. Reprocessing the data for improved and expanded statistical products will cost more than had we fully understood the data and processed the data more robustly in the beginning. It is the price of rapid prototyping.
  - The data are combined by an organization from several private sources. The data can be published because the data are anonymized, but methods used to obtain and process the data are covered by nondisclosure agreements. Looking under the hood to understand how to use the data and data quality has required negotiation. We are not even told the private sources of the data.
- Insights from one agency:
  - Most of the data providers furnish only aggregated data organized into granular “cells,” not true microdata. This limits the agencies’ ability to assess and use the data.
  - Most data series are available for only short periods of 2 or 3 years. This further limits the agencies’ ability to assess and use the data.
  - Many data series suffer from discontinuities, outliers, sample churn, and/or seasonality.
  - Documentation for both the underlying data and organization’s processing of it is incomplete.
  - From stakeholders, we heard some affirmation of our ongoing concerns. These include problems with sample composition and consistency of data content and access, among others.
- Agency to research best use of the data and then adapt it to our process. Largest challenge was the timing on our end due to busy schedule, limited resources, and other time sensitive pressures.
- [H]igher compute and storage requirements are necessary to store and process a constellation of over 140 satellites the cover the Earth daily. Traditional federal

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



computing facilities are insufficient to store and process daily national 3–5 meter resolution coverage.

- The data will be used for modeling and the creation of synthetic data. We have yet to start this process. Creating synthetic data is not something that the division has experience with. As such, we will hire contractors with this expertise to create the synthetic data file.
- Another data set:
  - Inability to fully validate the accuracy of the data received, no benchmarking data.
  - Unable to view private data modeling algorithms.

Agencies were asked to describe their practices in sharing private datasets for federal use beyond the initial reason for acquisition.

**Almost all agencies (18 of 19) share private data internally**, that is, within the agency. Agency approaches to private data sharing are diverse in nature. Most agencies (13 of 18) that share data identified more than one approach.

General survey: How does your agency share private data within the agency? (Check all that apply)	Number	Percent
As needed by other internal groups or persons	11	57.9
Only if there is a work-related need	9	47.4
Standardized process to share data within contractual/MOU limits	7	36.8
Dependent on any additional costs to share data internally	7	36.8
The project team managing the private data determines	7	36.8
Other (Specify)	7	36.8
Data sharing is not permitted	1	5.3
Number of respondents	19	

On the other hand, the case survey indicates, sharing private data outside the agency does not appear to be the norm. There was only one instance in which data were shared outside the acquiring agency, and even then, it was only within the parent Department.



## Sustainability of private data use—risks

The general survey asked each private-data-using ICSP agency: “Thinking about the next five years, are you concerned that access to private data that your agency uses will **not** be sustainable?”

- **Ten of 19 respondents said they are concerned.**
- **All but two principal federal statistical agencies using private data are concerned about sustainability.**

Specific concerns regarding sustainability included the following:

- **Private data vendor actions:**
  - Raise prices
  - Change product characteristics (for example, reduce quality, granularity)
  - Change methodology
  - Change business model
  - Participate in merger or acquisition (leading to one or more of the above actions)
  - Goes out of business
- **Government actions:**
  - Reduce funding available to purchase private data
  - Increase privacy/confidentiality restrictions

One agency noted that there may be ethical concerns about purchasing data for specific populations while balancing transparency and providing people and businesses with a meaningful choice with respect to how their data are obtained and used.

One agency suggests that differences in sustainability risk is a function of the centrality of the private dataset to the vendor’s business model. It perceives that a dataset considered part of a vendor’s core business offerings has less sustainability risk than one that is a supplemental revenue stream for a vendor.

In each of the 10 cases in which the private dataset had been assessed, the agency respondent saw 1 or more potential risks in relying on the private data.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



Case survey: Looking ahead, what potential risks does your organization see for relying on these private data, if any?	Number	Percent
Vendor consolidation	6	54.5
Vendor cancelling a data series	6	54.5
Other (Specify)	6	54.5
Vendor unwilling to sell to government	5	45.5
Number of respondents	11	

In text remarks, respondents for several cases indicated specific additional risks (see the box “Case Survey Respondents on Other Risks of Private Data Reliance”).

### Case Survey Respondents on Other Risks of Private Data Reliance

- The vendor might price the data prohibitively such that it is no longer affordable; the data might no longer be available, and an alternative solution would be needed; the vendor may not maintain the representativeness of the data.
- Any disruption to data quality or consistency. Use of the same data by third parties (or by the data sources themselves) to predict official statistics in advance.
- The data depend on public willingness to allow mobile device applications to track activity. If the public opts out, potential bias increases or the data just go away. We recognize that the data may not be a lasting source as cell phone and other local device users opt out of tracking.
- These data are often expensive, which could present challenges. Additionally, contractual obligations in obtaining these data could change, causing operations risks (changes in vendors, scheduled deliveries, formats, etc.)
- Vendors are continually bought/sold/consolidated, so data continuity is a concern. Pricing can be unsustainable without market competition. Continuity is our major concern.
- The introduction of an additional data source may introduce potential bias that is not present in the current data sources.



## **Best practices and lessons learned in the acquisition, preparation, analysis, and sustainability of private data**

Throughout both surveys, agencies offered specific, often detailed comments that describe successful adaptations and responses to various challenges in private data acquisition, preparation, analysis, and sustainability. Grouped by objective, excerpts are offered.

### *Smooth procurement process*

- Identify and engage stakeholders across the agency who require acquisition of the data.
- Anticipate procurement process challenges and address them as soon as possible (including educating procurement staff as appropriate).
- Work with data providers who have experience in fulfilling federal statistical standards, methodologies, and classifications.
- Check with partner agencies or department to see if an agreement or contract already exists for use or adaptation.
- Anticipate legal challenges and address them as soon as possible.
- Ensure vendor knows that proprietary secrets will be protected.

### *Cost-effective use of funding*

- Seek out other agencies that currently or could use the dataset to coordinate a government-wide or multiagency purchase or licensing agreement.
- Conversely, avoid duplicate agreements or purchasing of data.
- Avoid dependency on vendor's proprietary tools.

### *Data transparency and clarity*

- Seek to include terms in contract that vendor will provide full transparency of methodology and agency will protect vendor intellectual property.
- Require full data documentation and validation results
- Require that vendor communicate when it introduces any changes in data coverage, availability, methodology, and information technology (IT) support.
- Thoroughly learn of data limitations and challenges from vendor before signing contract. Use sample datasets to get a better idea ahead of time.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- Receive data documentation early in the process to start the planning of integrating the purchased data with the collected data. This will help with the lengthy integration and harmonization process.

## *Adequate technical capacity*

- Understand technical requirements as soon as feasible and set up appropriate technical infrastructure in response.
- Identify technical skills needed and see that staff can acquire them.
- Arrange for onboarding and special sworn status for members of vendor support team.
- Engage assistance of academic data scientists as useful.

## *Vendor reliability*

- In contract, specify vendor responsibilities for data quality control.
- Work with vendors with experience in ensuring that data products meet federal statistical standards.
- Have the vendor create and introduce a competent support team for extraction, IT, and analytics. In particular, have vendor clearly identify points of contact to answer questions that arise as agency writes specifications for processing, use, linkage, and storage.
- Establish recurring touch points with vendor to discuss and address identified deficiencies.

## *High data quality*

- Rely on more established private sector data sets and sources.
- Prioritize data sources for which substantial multi-year history is available.
- Establish specific dataset layout, coverage, and delivery timing precisely in the contract.
- Attach high value to obtaining access to true microdata, with identifiers if possible.
- Allow for a period of review time before having to purchase the data, which allows for the identification of any major data issues.
- Establish repeatable process for evaluating scope, coverage, and accuracy of the vendor data (against authoritative sources where possible).



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- Conduct a pilot to test and ensure that the coverage and quality of the data and its associated methodology are as expected.
- If vendor data are not easily matched to existing agency data, have vendor work with agency to create a crosswalk.
- In agency surveys, collect items that correlate with ones in private datasets to support imputation as necessary.

## *Cost-effective data analysis*

- Build prototypes faster upon acquisition
- Spend more time experimenting with similar datasets before the next crisis pushes us into premature production.
- Specify an algorithm to aggregate data to the proper level, review data to ensure that aggregation yields reasonable results, and revise algorithm to address instances when it yields unreasonable results.
- Develop an extrapolation algorithm to impute empty cells in private datasets.
- Adapt use to the limitations of the data, rather than trying to find further solutions to improve the data.

Interestingly, less than half of the respondents (9 of 19) expect to facilitate use of private data through greater reliance on written guidance regarding data acquisition, preparation, and use. This expectation seems somewhat correlated with extent of reliance on private data. Of the seven agencies that extensively or moderately rely on private data, five expect to increase their use of written guidance.

## **Spinoff benefits**

Case survey respondents were asked to describe the nature of any benefits resulting from private dataset use beyond the immediate reason for acquisition. Seven respondents provided detailed positive observations (see the box “The Added Benefits of Private Data Beyond Initial Use”).



## The Added Benefits of Private Data Beyond Initial Use

- Case #1
  - This experience demonstrates the power of having a source of “nearly” real time information for purposes of complementing other data used to produce official statistics. While one cannot be certain that future procurement of private data can provide such impact, this experience reinforces the notion that alternative datasets, including private data, will continue to be important in meeting the mission.
  - Our initial article led to input/feedback that helped us improve the methodology and deeper understanding of the data, benefitting the expanded use during the pandemic by allowing us to quickly execute and adjust data use.
- Case #2
  - As a result of this project, agency identified ways to streamline the research and approval processes for utilizing private data sources of price data.
  - During the development of the index methodology, agency incorporated improvements based on expert review.
  - Agency is trying to replicate this model with other sources of private data that are available for purchase. Agency recognizes the benefits from expanding the size [number] of transactions priced for certain product areas.
- Case #3
  - Agency launched a minimally viable statistical product quickly without fully understanding the data to meet demands of senior officials for statistics yesterday. As we have explored the results, we see new ways to process the data for additional and improved products. Reprocessing the data for improved and expanded statistical products will cost more than had we fully understood the data and processed the data more robustly in the beginning. It is the price of rapid prototyping.
  - Expanding the range of measures extracted from the dataset to capture travel through and to locations as well as from locations.
  - The data were developed with research funds from another agency. We are coordinating with the original sponsor to assure that applications by both agencies use consistent definitions, etc.
- Case #4
  - We are considering the potential for “off-label” uses. For example, credit card transaction data have been widely used to study consumer spending. Could these



data also be used to measure merchant births and deaths as an input to estimating changes in employment?

- Case #5
  - The use of these data may set a precedent and encourage broader uses of these types of data in the agency in the future.
- Case #6
  - The data are shareable within the agency Google Cloud Platform environment. Other sister agencies are exploring the utility of this data stream.
  - Agency wants to continue exploring new space-based initiatives that can contribute to improving ag statistics. The data's daily high-resolution collections provide a new opportunity to determine if this platform can improve monitoring for crop area and disaster assessments.
- Case #7
  - These data are being sought by agency research and program offices for operational planning analysis.

## Conclusions

With the implementation of the Evidence Act, the Interagency Council on Statistical Policy expanded from 13 to 26 members. Within 6 months of that expansion, the COVID–19 pandemic put a hiatus on in-person meetings and led to major disruptions in various statistical activities. Together, this series of events brought to light the expanded need for ICSP members to learn from each other, share best practices, and work together toward new approaches to obtaining, processing, and disseminating information. This exercise on private sector data is a positive example of how the ICSP agencies can share information designed to improve future statistical processes.

The broad takeaways from this activity are the following three: (1) the widespread use of private sector data across the statistical system, (2) the array of challenges experienced in the process, and (3) the capacity of agencies to swiftly move themselves up the “learning curve” in response to these challenges.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



- **ICSP member agencies have embraced private sector data as a part of their toolkit—and offer multiple reasons for doing so.** Nearly all responding ICSP agencies indicated the recent use of private sector data, and some identified extensive use. Such data were frequently used to supplement or combine with existing data or to provide context or validation. These data may also be helpful in addressing the interest in high-frequency data. There were practical reasons as well, including ease of obtaining private sector data and lower cost compared to direct data collection.
- **Agencies identified challenges in acquiring, preparing, analyzing, and sustaining the use of private data.** Among concerns identified with current use included high cost, legal hurdles, lack of granularity, and uncertain quality. Concerns were also raised about future private sector data use, where agencies worried about continued availability, rising costs, consistent quality, and related variables. Perhaps these concerns are best summed up as a need to be diligent in the assessment of private sector data. Many agencies noted the considerable time and care they took before making the decision to use private sector data.
- **In response to these challenges, agencies have generated—and are ready to share—a growing set of lessons learned and best practices in private data use.** The working group is impressed with the efficacy and flexible adaptability of various agency efforts and hopes that ICSP members find the details offered of value.

Falling survey response rates, greater access to administrative records, and the desire for more information drives the statistical agencies to seek alternatives to traditional data collection. The examples identified in this report show that many of these alternatives have been successful and require agencies to learn new methods, techniques, and skills. Hopefully, the lessons identified in this report will aid the efforts of federal statistical leaders as they seek new data horizons.

In addition, the working group believes, this project demonstrates how collaborative ICSP efforts can aid each member agency in achieving its mission. While agency staff must attend to production responsibilities before they can turn to ICSP coordination activities, the value of this report's survey findings makes clear that a modest contribution from each agency can lead to robust results for all.



# Appendix: Survey Instruments

Interagency Council on Statistical Policy  
Private Sector Data Project

## General Survey

This survey is sponsored by the Interagency Council on Statistical Policy (ICSP) through the joint efforts of federal statistical agencies and the George Washington Institute of Public Policy (GWIPP). We anticipate this survey to take about 15 minutes of your time. Please provide your response within two weeks <BY DATE>. This survey is administered by a third party that will have access to the information provided. DO NOT include any confidential data in your answers.

The purpose of this research is to better understand the breadth and nature of federal statistical agency uses of private sector datasets. In this survey, **private sector data** is defined as:

- ***Data purchased or licensed from private-sector entities within the past five years.***

When thinking about private sector data, exclude the following:

- Data traditionally collected by statistical agencies directly from private companies, for example, through surveys.
- Data extracted or collected solely from public administrative records.
- Data scraped from the web that is not purchased or licensed from a private company.
- Data that is freely accessible from public companies or websites.
- Data purchased or licensed more than five years ago.

Please focus on **statistical or data operations** at your agency.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



1. Has your agency purchased or licensed private data from private-sector entities within the past five years?

- Yes – continue to question 2
- No – Check all of the following reasons that apply:
  - Data are too expensive
  - Data meeting agency needs is not available
  - Available data do not meet agency quality standards
  - Difficulty in getting contracts approved
  - Agency does not have a need for private-sector data
  - Other (Specify)

*Please specify here*

If you answered NO to question 1, END OF SURVEY.

2. Over the past five years, for what reasons has your agency used private data?

[Check all that apply]

- To make available to agency stakeholders and data users (publish data to the web)
- To continue current reporting capacity of agency priorities
- To supplement or combine with existing agency held data
- Verification, quality control or quality assurance for existing data or estimates
- For use as or to identify a survey frame
- To better understand other indicators of the economic environment (situational awareness)
- Direct data collection is too costly (private data sets are a cost-efficient alternative)
- Direct data collection is too burdensome on respondents (resulting in poor response or data quality)
  - Other (Specify)

*Please specify here*

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



3.a. Over the past five years, what were any challenges in **obtaining** private data?

Check here if there have been no challenges over the past five years.

Otherwise, check all that apply:

- Legal hurdles
- Cost
- Difficulty obtaining sufficiently granular data
- Technological challenges in accessing data once obtained
- Security challenges in accessing data once obtained
- Other (Specify)

*Please specify here*

3.b. Please describe any solutions or best practices to address the challenges you identified.

*Please describe here*

4.a. Over the past five years, what were any challenges in **analyzing or using** private data once obtained?

Check here if there have been no challenges over the past five years.

Otherwise, check all that apply:

- Data are of poor quality or incomplete
- Data require cleaning or additional manipulation to make useful
- Data are not easily matched to existing agency data
- Data do not fully capture or match agency measurement objectives (construct validity)
- Data are not fully representative of the measurement universe (agency frame)
- Methodology is not clearly documented (collection or aggregation methods are not clear)
- Other (Specify)

*Please specify here*

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



4.b. Please describe any solutions or best practices to address the challenges you identified.

*Please describe here*

5.a. Thinking about the next five years, are you concerned that access to private data that your agency uses will **not** be sustainable?

- Yes
- No

5.b. If yes, describe any specific concerns.

*Please describe here*

6. What are your agency's approach(es) to managing private data contracts or memorandums of understanding (MOUs) for private data?

[Check all that apply]

- Centrally (at the agency level)
- Decentralized (at sub-agency level, such as individual program)
- Varies by project
- Limited to a specified dollar amount
- Other (Specify)

*Please specify here*



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



7. How does your agency share private data within the agency?

Check here if data sharing is not permitted.

Otherwise, check all that apply

- Standardized process to share data within contractual/MOU limits
- As needed by other internal groups or persons
- The project team managing the private data determines
- Dependent on any additional costs to share data internally
- Only if there is a work-related need
- Other (Specify)

*Please specify here*

We are interested in understanding the extent to which your agency currently uses private data.

8.a. How many **statistical programs** does your agency publish? *Count each statistical program once, such as the Consumer Price Index or the National Income and Product Accounts.*

|\_|\_|\_|

8.b. Roughly speaking, how many of your agency's **statistical programs** use at least some private data? *Include any programs that solely use private data.*

|\_|\_|\_|

8.c. Roughly speaking, how many **private data sources** does your agency use in production?

|\_|\_|\_|



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



9.a. Has your agency adopted a set of principles or criteria regarding the use of private data?

- Yes
- No
- Unsure

9.b. If yes, would you be willing to share it?

- Yes
- No

10.a. Has your agency specified the use of any private data in current budget justification, strategy, plan, and/or performance documents?

- Yes
- No

10.b. If yes, please briefly describe or provide linked citations.

*Please describe here*

11.a. In the last five years, has your agency sought to obtain private data that is outside of any budget initiative, strategy, or plan in response to a specific opportunity or need?

- Yes
- No

11.b. If yes, please briefly describe.

*Please describe here*



12a. Looking ahead, does your agency expect to facilitate use of private data through greater reliance on written guidance?

- Yes
- No

12.b. If yes, please briefly describe.

*Please describe here*

13. Please briefly describe up to three cases of your agency’s experiences with current or planned uses of private data. Possible cases, for example, might include a private dataset with a high impact on your organization; a private dataset project that has not yet come to fruition due to significant challenges; or a project currently in development you’re excited about. We suggest one paragraph per case, with a sentence or two on each of:

- data source/intended uses;
- catalyst for use;
- outputs/outcomes;
- and experience/lessons learned.

We intend to follow up with some additional questions regarding one or more of the examples you provide.

1.

2.

3.

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



## Interagency Council on Statistical Policy Private Sector Data Project

### Case Survey

In the first ICSP Private Sector Data Survey, your organization identified the following private data project/dataset/source:

....

Please answer the following questions thinking about this use of private data. We anticipate this survey to take about 15 minutes of your time. Please provide your response within two weeks (Tuesday, August 31<sup>st</sup>). This survey is administered by a third party that will have access to the information provided. DO NOT include any confidential data in your answers.

#### Section 1: Data use

1. Describe how these private data are or will be used at your organization [Check all that apply]:

- To make available to agency stakeholders and data users (publish data to the web)
- To continue current reporting capacity of agency priorities
- To supplement or combine with existing agency held data
- Verification, quality control or quality assurance for existing data or estimates
- For use as, or to identify, a survey frame
- To better understand other indicators of the economic environment (situational awareness)
- In lieu of direct data collection, which is too costly (private data are a cost-efficient alternative)
- In lieu of direct data collection, which is too burdensome on respondents (resulting in poor response or data quality)
- Other (Specify)

*Please specify here*

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



2. What is it about these private data that prompted your organization to highlight it for the ICSP?

*Please describe here*

3. a. Did your organization's use, or planned use, of these private data change beyond the original scope of work?

- Yes  
 No

- b. If yes, briefly describe how this use changed beyond the original scope of work.

*Please describe here*

4. a. Has your organization shared information on your experience with obtaining and using these private data?

- Yes  
 No

- b. If yes, briefly describe any improvements you made based on sharing with others.

*Please describe here*

- c. If yes, briefly describe any concerns that were identified based on sharing with others.

*Please describe here*



5. a. Has your organization shared these private data with other federal agencies?

- Yes
- No
- Sharing is not allowed
- Not applicable – project still in the planning stage

b. If yes, describe what made sharing these private data successful, and any barriers your organization encountered.

*Please describe here*

6. a. Has your experience with these private data affected how your organization approaches the use of private data now or in the future?

- Yes
- No
- Not applicable – project still in the planning stage

b. If yes, briefly describe.

*Please describe here*

7. To inform best practices, what additional challenges, barriers, or missteps would your organization like to share on the use of these private data?

*Please specify here*



## Section 2: Acquisition and Assessment

1. Did your organization conduct any assessment of data quality before acquiring these private data?

- Yes  
 No

If yes, briefly describe the quality assessment that was conducted.

*Please describe here*

2. From what organization(s) were these private data obtained? If confidential, indicate "confidential."

*Please describe here*

3. How were the private data obtained? [Check only one response]:

- Obtained through an agreement (memorandum of understanding, contract, or other agreement)  
 Obtained informally (such as web scraping or downloading)  
 Other (Specify)

*Please specify here*



4. For private data that were obtained through an agreement, were data “off-the-shelf” or customized?

- Off-the-shelf
- Customized
- Other (Specify)

*Please specify here*

5. a. Once acquired, how did/does your organization assess data quality? Check all that apply.

- Formal quality assessment methods like [https://nces.ed.gov/fcsm/pdf/FCSM.20.04\\_A\\_Framework\\_for\\_Data\\_Quality.pdf](https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf)
- Availability of detail of provider's methodology
- Comparability to existing data
- Compatibility to traditional official statistics
- Not applicable because we conducted a detailed assessment before purchase
- Other (Specify)

*Please specify here*

b. Would your organization be willing to share any written process or model you have to assess quality?

- Yes
- No



# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



6. This question attempts to learn about the resources your organization expended on these private data.

First, we are interested in the financial cost to your organization for acquiring and maintaining these private data.

- a. What was the initial cost to acquire private data from the organization providing it? If another organization was not involved, for example, if these data were freely scraped from the web, report \$0. DO NOT include costs associated with staff time or equipment (for example, server space).

\$ \_\_\_\_\_

- b. What is the annual license or maintenance cost for continued use of these data?

\$ \_\_\_\_\_

7. Next, we are interested in how easy or difficult it was for staff or contractors to prepare and use these data? *Using a scale of 1 (very easy) to 5 (very difficult)*

- a. How would you rate how easy or difficult it was for staff or contractors to get these data to a point where analysts could review or test these data.

Choose an item.

- b. If these data are in production, how would you rate how easy or difficult it is for staff or contractors to use these data each time?

Choose an item.

8. How many months did it take from data acquisition until data were in suitable shape for analysis?

Click or tap here to enter text.

9. a. How many months did it take from suitable shape for analysis until data were used in production?

- b. Describe briefly any challenges experienced in preparing these data for use after acquiring.

*Please describe here*



10. a. What terms and conditions, if any, were imposed by the private data provider?

- No terms and conditions; *otherwise Check all that apply*
- Term licensing
- Perpetual ownership
- Republication rights [verbatim, aggregated, only images of a chart, none]
- Site-wide licenses to reduce copyright infringement within the organization
- Inability to publicize vendor name
- Requirement to publicize vendor name
- Other (Specify)

*Please specify here*

b. Please elaborate on any unique terms and conditions.

*Please elaborate here*

11. Looking ahead, what potential risks does your organization see for relying on these private data, if any?

- Vendor consolidation
- Vendor cancelling a data series
- Vendor unwilling to sell to government
- Other (Specify)

*Please specify here*

# The Use of Private Datasets by Federal Statistical Programs: Extent, Challenges, and Lessons Learned

January 6, 2023



12. What would your organization do differently on future acquisitions/assessments in light of your experience with this example?

*Please describe here*

13. Please include the name and email of a point of contact from your agency in the event we would like to follow up regarding any survey responses. Thank you.

*Please enter here*