



The Congressional Budget Office's Access to Data From Federal Agencies

JUNE | 2021

During consideration of the fiscal year 2021 appropriation bill for the legislative branch, the House Committee on Appropriations requested information about the Congressional Budget Office's access to data from federal agencies, including data sources and data sets. This report provides that information.

To fulfill its mission, CBO accesses a wide array of data from federal agencies. CBO uses those data for producing baseline budget projections, economic projections, cost estimates, and reports. The Congressional Budget Act of 1974 (the Budget Act) provides CBO general authority to access data from a variety of sources. CBO also accesses data by using specific authority or by collaborating with other agencies.

CBO uses both public data and data that are not public (because they are confidential, proprietary, or otherwise restricted). Using publicly available data can help the agency respond to Congressional requests in a timely manner, but those data do not always provide the information needed for the level of analysis requested. Sometimes additional information is available from the agencies even if it has not been released to the public. When public data are insufficient to answer a question, CBO's analysts can obtain data from agencies through either informal or formal agreements. CBO currently has more than 20 active data use agreements with other federal agencies.

Restricted data, and especially restricted data linked to other sources (known as commingled data), tend to offer more information and are useful for examining a broader set of issues of interest to the Congress. However, access to restricted data, and the release of analytical products

based on those data, may be hindered because CBO must navigate multiple legal authorities and ensure that the data remain secure throughout the process.

CBO accesses restricted data in a variety of ways. The most common way for CBO to access such data is for the owners to transmit the data to CBO to be housed on-site or in an approved system in the cloud (an Internet-based environment supplying computing, storage, and software infrastructure as a service). That involves following specified security procedures, which vary depending on the agency and the data. CBO may also access restricted data through another agency's systems or through a third party, such as a contractor.

CBO is obligated to protect data in the same way that other federal agencies do. When data are collected for statistical purposes, they must be protected to ensure confidentiality. Enhancements in computing power and the increasing availability of outside data sources are leading to changes in standards and best practices for maintaining privacy, including the implementation of formal privacy methods. Those changes may increase the time needed for some analyses and thus affect CBO's ability to be responsive to the Congress.

Legal Background

CBO has good working relationships with most executive branch agencies and can often obtain information simply by asking for it. That is, CBO frequently relies on cooperation among the branches of government. Sometimes, however, formal arrangements are required. In those instances, three different types of legal authority help CBO obtain information: *general* authority, which requires agencies to provide information to CBO; *specific* authority, which requires agencies to provide particular

information to CBO; and *other* authority, which permits agencies to share information with CBO as needed for CBO to perform work that benefits those agencies.

General Authority

The Budget Act—specifically, section 201(d)—authorizes CBO to obtain “information, data, estimates, and statistics” from executive branch agencies.¹ It also requires executive branch agencies to provide CBO with any material that the Director of CBO determines to be necessary for the performance of the agency’s duties and functions “other than material the disclosure of which would be a violation of law.”² As part of the annual budget process, federal agencies are also required to provide CBO with “fiscal, budget, and program information.”³ General authority is also provided by the Federal Credit Reform Act of 1990, which stipulates that CBO “shall have access to all agency data that may facilitate the development and improvement of estimates of costs” of federal direct loan and loan guarantee programs.⁴

Specific Authority

CBO also has specific authority to obtain certain data. For example, the Internal Revenue Code authorizes CBO to use federal tax information for long-term models of the Social Security and Medicare programs.⁵ The Department of the Treasury is required to make available to CBO information related to the Troubled Asset Relief Program.⁶ CBO is authorized to receive price information about drugs purchased by the Department of Veterans Affairs (VA), the Department of Defense,

the Public Health Service, and the Coast Guard.⁷ CBO can also receive detailed information about the rebates and discounts that Medicare Part D plans and their pharmacy benefit managers receive on their purchases of prescription drugs.⁸

Other Authority

CBO is granted access to some information from other agencies because that information advances the other agency’s mission. For example, the Chief of Staff of the Joint Committee on Taxation (JCT) designates certain employees of CBO as his or her agents to inspect tax returns and return information (without identifying individual taxpayers).⁹ JCT uses CBO’s baseline projections—which are developed in collaboration with JCT—when performing its revenue estimation responsibilities. It is useful for the two agencies to be able to share common modeling assumptions, techniques, and data to enhance the accuracy and consistency of their respective estimates and for CBO to prepare accurate and consistent baseline estimates that JCT can use as a basis for its revenue estimates.

Similarly, some of CBO’s employees have access to confidential data from the Census Bureau because their work has the potential to benefit the Census Bureau’s programs and activities. Those CBO employees are sworn to use census data only for statistical purposes and to protect the confidentiality of those data.¹⁰ CBO’s employees must identify and document how each project is anticipated to benefit the Census Bureau.

Other agencies can be assured that regardless of the authority used, CBO will protect the information it receives. Section 203(e) of the Budget Act requires CBO to maintain the same level of confidentiality as the agency from which the information is obtained. CBO’s employees are subject to the same penalties for

1. Codified at 2 U.S.C. §601(d) (2018). A similar provision for obtaining information from legislative branch agencies is found in section 201(e) of the Congressional Budget Act of 1974, Public Law 93-344 (codified at 2 U.S.C. §601(e) (2018)).

2. Sec. 201(d) of the Congressional Budget Act of 1974, P.L. 93-344 (codified at 2 U.S.C. §601(d) (2018)).

3. 31 U.S.C. §1113(b) (2018).

4. Sec. 503(d) of the Congressional Budget Act of 1974 (codified at 2 U.S.C. §661b(d) (2018)).

5. 26 U.S.C. §6103(j)(6) (2018).

6. Sec. 201 of the Emergency Economic Stabilization Act of 2008, P.L. 110-343 (codified at 12 U.S.C. §5251 (2018)). The information available to CBO comprises all information and records that the Government Accountability Office is entitled to review under section 116(a)(2)(C) (codified at 12 U.S.C. §5226(a)(2)(C) (2018)).

7. Sec. 1927(b)(3)(D) of the Social Security Act (codified at 38 U.S.C. §8126(e)(4) (2018)).

8. Sec. 1150A(c)(3) of the Social Security Act (codified at 42 U.S.C. §1320b-23(c)(3) (2018)); sec. 1834A(a)(10)(C) of the Social Security Act (codified at 42 U.S.C. §1395m-1(a)(10)(C) (2018)); and sec. 9114(a) of the Consolidated Omnibus Budget Reconciliation Act of 1985, P.L. 99-272 (codified at 42 U.S.C. §1395ww note (2018) (Allowing the Secretary of Health and Human Services to share information regarding payments to hospitals under section 1886 of the Social Security Act)).

9. 26 U.S.C. §6103(f)(4) (2018).

10. 13 U.S.C. §23(c) (2018).

unauthorized disclosure or use as employees of that agency.¹¹

Sources of Data

CBO uses various types of data from many agencies across the federal government. This report is not intended to be an inventory of data or data sources. Some of the agencies that CBO works with most frequently are within the Departments of Agriculture, Commerce, Defense, Education, Energy, Health and Human Services (HHS), Homeland Security (DHS), Labor, the Treasury, and Veterans Affairs.

CBO uses many forms of data from other agencies, and some of those data can be categorized in more than one way. Examples of the types of data the agency uses include the following:

- **Tabulated Data.** Many agencies, especially statistical agencies and offices, disseminate data in tabular or summary form. Those data may be released on a regular basis, such as annually or monthly, as a series.
- **Microdata.** Microdata are unit-level data from surveys, censuses, or administrative records. Examples include data collected about people or firms in a survey, data about each person or household in a census, and data collected on a tax return about each person in the tax-filing unit or information on the entire tax-filing unit. (A tax-filing unit is a single person or a married couple plus any dependents.)
- **Restricted Data.** Access to certain data is restricted. Sometimes CBO's access to such data occurs informally through conversations between analysts or through an agency's legislative liaison, while at other times a formal data use agreement is required.
- **Commingled Data.** Commingled data are data sets that are merged from multiple sources. Microdata can be commingled with other microdata at the same or more aggregated unit levels of observation or with tabulated data.

Tabulated data are typically publicly available but may be acquired through informal agreements with other agencies. Microdata can be publicly available or restricted. Typically, accessing restricted microdata requires a formal agreement. Commingled data take many forms,

but there are special considerations when one or more sources consist of restricted microdata.

Publicly Available Data

Publicly available data typically come in two forms: tabulated data and microdata. Using publicly available data allows CBO's analysts to respond to the needs of the Congress in a timely manner. Although analysts must be familiar with the data in order to analyze the information quickly, publicly available data are generally easy to access and are usually well-documented. Those data do not contain information about particular individuals or businesses.

Tabulated Data. CBO uses publicly available, tabulated data series for many purposes. Those data are generally easy for CBO's analysts to access quickly and are well-documented. In addition, the limitations of such data are well-understood. Challenges can arise when there is a break in the series, which happens when there are major changes in methodology, data collection, definition standards, or a lapse in the other agency's funding.

Examples of publicly available tabulated data series that CBO's analysts use regularly include monthly unemployment rates and price indicators from the Bureau of Labor Statistics (BLS), national income and product accounts from the Bureau of Economic Analysis, and data from the Census Bureau, such as industry indicators. Those types of data tend to be released on a regular schedule, so updates are predictable, and CBO's ability to respond to Congressional requests can be planned.

Microdata. Sometimes statistical offices and agencies also disseminate public-use versions of the microdata underlying a tabulated data series, allowing CBO to conduct independent analysis of the data. In almost all instances, before a file is released, the information collected undergoes processes to protect the confidentiality of the people or institutions that provide the information.¹²

CBO uses microdata from many different sources. Examples of publicly available microdata that CBO's analysts use on a regular basis include the American Community Survey, the Current Population Survey and

11. Codified at 2 U.S.C. §603(e) (2018).

12. Those processes can involve removing some information from the file, aggregating information, or "perturbing" reported values. (Perturbing values entails adding a small amount of random error to keep the reported value private without altering the overall distribution.)

its supplements, the Medical Expenditure Panel Survey, the National Health Interview Survey, the Statistics of Income Public-Use Microdata Files, and the Survey of Income and Program Participation.¹³ Those data come from the Census Bureau, BLS, the Agency for Healthcare Research and Quality, the Centers for Disease Control and Prevention, and the Internal Revenue Service (IRS). Although those data are available directly from the agencies that produce them, CBO sometimes obtains such data from other institutions that work to harmonize information collected at different times or to create tools to make analysis easier. Examples include IPUMS (which, before 2016, stood for Integrated Public Use Microdata Series) and the Inter-University Consortium for Political and Social Research (ICPSR).

There are limitations to publicly available data. Nuances to the data sometimes take time to understand. Also, agencies that produce data change their data collection and processes over time to adapt to new policies, new interests, and new circumstances. Although those agencies strive to improve the data, improvement necessarily means change. That change can make data less comparable over time and can require additional work on the part of analysts using the data. Finally, publicly available data may have less detail than underlying source data.

Data Acquired Through Informal Cooperation

When public data are insufficient to answer a question, CBO's analysts can obtain data from agencies in two ways—through informal agreements and formal agreements. Informal agreements can include qualitative or quantitative data. Those data may not be released publicly because they are sensitive, in which case they are kept for internal use only, or because of resource constraints at the other agency.

CBO's analysts regularly communicate with employees at other agencies about how a proposed policy might be implemented or to acquire additional information that might not be obvious from the data that are disseminated publicly. Several elements affect how responsive other agencies are to CBO, and in turn how responsive CBO can be to the Congress. Those include the relationships between analysts at CBO and the other agency, the interest of the other agency in responding to the Congress, and the other agency's resource constraints.

Some agencies give quantitative data informally to CBO's analysts as well, mostly in the form of summary statistics or tabulated data. Sometimes those are regular updates that CBO uses in its models, such as an agency's own projections. Examples include VA's projection of enrollment in its programs and health care utilization, additional information underlying population projections from the Social Security Administration (SSA), and additional information on new lawful permanent residents and approved applications from DHS.

Obtaining data from agencies can be challenging. Those challenges usually take the form of multiple layers of review and approval within an agency before the information is shared, which can cause delays. That type of delay is more common if it is the first or only time that the information has been requested, or if the request is tied to estimating the cost of a policy proposal that the other agency does not support. Both qualitative and quantitative data can be delayed because of reviews and approvals, making it more difficult for CBO to respond to the Congress in a timely manner. However, those processes may increase the accuracy of the information and, in turn, the accuracy of CBO's estimates.

Restricted Data Acquired Through Formal Agreements

Agencies cannot release all data that they collect to the public. Some of the information they collect is sensitive and releasing that information could harm the people they intended to help, endanger national security, or reduce public trust in those agencies. However, sometimes additional data are needed for CBO's work. Therefore, CBO's analysts rely on the legal authorities listed above to use data that are restricted, doing so in a way that protects the data that may not be released publicly.¹⁴

CBO's analysts typically access restricted data after a formal agreement has been made between CBO and the other agency. Those agreements are usually memoranda of understanding, and they require input from analysts and attorneys from both agencies to ensure that specific data needs are met and that legal requirements

13. CBO also uses the restricted version of the Statistics of Income for certain purposes as agents of JCT.

14. Common examples include title 13, which governs the confidentiality of data collected by the Census Bureau; title 26, which governs the confidentiality of tax data; and the Confidential Information Protection and Statistical Efficiency Act of 2002, which applies to data collected by other statistical agencies and offices.

are satisfied. There are additional considerations when restricted data come from multiple sources.

Data Agreements. Formal data agreements are beneficial because they fully lay out the scope, responsibilities, and expected output of CBO and the agency providing the data. Most data use agreements pertain to accessing microdata. Such agreements typically specify how CBO's analysts should access the data and which analysts are allowed access. They also stipulate requirements for physical and electronic security, requirements in case of a breach of security or unauthorized access, and processes for publishing information based on the data. Some agreements also specify what topics CBO's analysts can study with the data. Examples of restricted data accessed under agreements include confidential data from the Centers for Medicare & Medicaid Services (CMS) about rebates on Medicaid and Medicare Part D drug purchases, which include proprietary business data for drug manufacturers and Part D plans; data from HHS about Temporary Assistance for Needy Families; and data from the Department of Education on federal student aid applications.

CBO currently has more than 20 active data use agreements with other federal agencies. Those agreements typically allow CBO to access data for three to five years if an expiration date is specified. The time required to develop and finalize a data use agreement can range from 5 to 60 hours for CBO's analysts, attorneys, and other personnel; however, the total amount of time that elapses before an agreement is reached can be anywhere from less than one month to, in extenuating circumstances, up to five years. Time may also be needed for training, reporting, audits, and renewals. The total amount of time required for an agreement to be finalized also depends in part on the responsiveness of the other agency or agencies. The time needed to formulate and finalize an agreement limits the types of projects for which such agreements are useful. Therefore, restricted data are used more often in longer-term and recurring projects than in projects with shorter deadlines.

Typically, data use agreements cover microdata that may include sensitive information; therefore, CBO is required to maintain the same level of confidentiality as the agency that provides the data, and CBO's employees are subject to the same rules as the other agency's employees. Also, the authority governing the collection or sharing of data may specify how the data can be used, which is then

specified in the formal agreement. For example, through direct authority from title 26, tax records are shared with CBO through SSA for the purpose of long-term modeling of Social Security and Medicare.¹⁵ Broadening allowed uses for CBO could help the agency provide answers to the Congress more quickly and with more depth.

The Foundations for Evidence-Based Policymaking Act of 2018 (the Evidence Act) is meant to increase accessibility to federal data for policymaking and evaluation, but that law has not been fully implemented.¹⁶ The act requires agency data to be accessible and requires agencies to plan to develop statistical evidence to support policymaking.¹⁷ Full implementation may result in some improvements in the timeliness of CBO's access to data. In particular, the Office of Management and Budget is planning to standardize the application process for accessing restricted data. That standardized application could make the process easier for CBO's employees and the review by other agencies more expeditious. Also, standardized guidance on privacy protection is expected to be issued.

Whether the Evidence Act will lead to changes in the data CBO can access or in the time frame the agency needs to be responsive to the Congress remains to be seen. Like the Budget Act, the Evidence Act also exempts data that have restricted uses under other laws. It is unclear whether the Evidence Act will broaden CBO's access to data relative to other authority the agency already had before the act's implementation.

Commingled Data. Commingled data are data from multiple sources that are combined. Analysis can be enhanced by using multiple sources to get the best available data on different topics and including information that is not available in a single data set.

If one source is publicly available and the other source is restricted, then the data must be protected under the guidelines established for the restricted data source.

15. 26 U.S.C. §6103(j)(6) (2018).

16. The implementation schedule is currently delayed. See Government Accountability Office, *Open Data: Agencies Need Guidance to Establish Comprehensive Data Inventories; Information on Their Progress Is Limited*, GAO-21-29 (October 2020), www.gao.gov/assets/gao-21-29.pdf (1.43 MB).

17. The Foundations for Evidence-Based Policymaking Act of 2018, P.L. 115-435, 132 Stat. 5529 (2019). See <https://go.usa.gov/xHFrfq>.

Sometimes, although not always, adding publicly available data to restricted data can increase the risk of “reidentification.” For example, merging information about local economic conditions with a person’s response to a survey could increase the risk that more precise geographical information about the respondent could be inferred from resulting publications; that risk would need to be mitigated.

When commingled data come from multiple restricted sources, particular challenges and opportunities can arise. Challenges stem from issues related to privacy and security. Ensuring that the data are secure means creating agreements that include multiple agencies, navigating different data collection and use authorities, and protecting the data under multiple authorities.

Creating Agreements. Writing agreements to use commingled data can be complex because of the different authorities that govern the collection and use of data. For example, the Census Bureau can access tax data for statistical purposes only, and those data are also available to other researchers under certain conditions. The Evidence Act defines statistical purposes as the “description, estimation, or analysis of the characteristics of groups,” along with other purposes.¹⁸ Tax data and other administrative data commingled with survey data could be useful for CBO’s work, enabling it to improve baseline projections, estimate distributional effects of legislative proposals, and better estimate the effects of changes to Medicare. However, the Department of Commerce’s ability to access tax data for statistical uses has been more narrowly interpreted than what is described under the Evidence Act.¹⁹ It is unlikely that the more broadly defined statistical uses described under the Evidence Act will also be considered as statistical uses under the Census Bureau’s authority to access tax data.

Because of the complexity of agreements governing access to commingled restricted data, CBO’s ability to use such data in a time frame that allows the agency to be responsive to the Congress can be hampered. The issue has particular relevance for requests for data detailing the distributional impacts of policies by demographic characteristics—such as race, ethnicity, or disability

status—that are not collected by the IRS for tax-filing purposes. For example, for close to a year, CBO has been working with the Census Bureau and the IRS to create an agreement that would allow CBO to have broader, ongoing access to commingled survey and tax data, but an agreement still has not been signed. Currently, each project that uses commingled data from the Census Bureau is approved individually, and CBO’s analysts go through the same process that academic researchers do; approval of individual projects takes at least four months, and usually more. The time line depends on the complexity of the project and the time needed to fully develop a proposal. Only after that approval is obtained can CBO begin the days, weeks, or months of work needed to complete the data analysis and publish the findings.

Creating agreements with multiple agencies is particularly difficult because of the amount of resources involved and the larger number of people whose interests would need to align. For example, in order to access data from the Health and Retirement Survey linked to data from CMS, agreement is required between CBO, CMS, the National Institute on Aging, Acumen, and the University of Michigan.²⁰ In many cases, CBO applies to use data following the same processes used by academic researchers; such an approach does not always allow the agency to be responsive to the Congress.

Further, some agencies have concerns other than legal constraints. BLS has declined to approve a project with commingled data because they “assure respondents that their data are kept strictly confidential, and [they were] concerned that such research would present an appearance issue.” Many topics that CBO works on are sensitive in nature because they are of interest in the Congress; consequently, that type of restriction can be difficult to reconcile with CBO’s data needs.

Using Commingled Data. When commingled data come from multiple restricted sources, there are more risks of unauthorized disclosures. Data that come from more than one source are protected by the authority that covers each data set. For example, survey data collected by the Census Bureau under title 13 commingled with

18. 44 U.S.C. §3561(12) (2018).

19. See Enrique Lamas, Chair, Data Stewardship Executive Committee, “DS002: Policy on Title 13 Benefit Statements” (Census Bureau, May 22, 2018), www2.census.gov/foia/ds_policies/ds002.pdf (156 KB).

20. The University of Michigan produces the Health and Retirement Survey, and the National Institute on Aging is the custodian for CMS-linked data. Those linked data require storage in a secure enclave (a network that holds confidential data), which is hosted by Acumen.

data from the IRS under title 26 are subject to protections from both laws, and analysts would face penalties under both laws for unlawful disclosure. Analysts need to obtain special sworn status to access those data.

Although obtaining permission to access those data is time consuming, data that come from multiple restricted sources and are linked together tend to be especially rich and allow CBO to answer questions it otherwise could not. Because different data sources have different strengths, combining data from multiple sources means that the analyst can decrease measurement error and uncertainty by using the best available information from each source. Commingled data also offer more information than any of the individual sources, which allows for better quality analysis and less uncertainty in the results.

Accessing Data

Data can be accessed in a number of different ways. Public data are typically downloaded from agencies' websites and saved on CBO's network drives. Restricted data can be transmitted to CBO and accessed within the agency's own information technology (IT) infrastructure, accessed through the data owners' infrastructure and systems, or accessed through a third party. CBO is open to all such arrangements to facilitate the needs of each agency. In particular, CBO respects the resource constraints of other agencies and therefore seeks to be flexible in its approach.

Accessing Restricted Data Through CBO's Infrastructure

The most common type of agreement permitting access to restricted data involves transmitting the data to CBO. The data are then used and maintained under the agency's purview and under the specified terms of the agreement.

Securely transmitting restricted data to CBO must be done in a way that protects the data as required by law. That is, just as the data must be stored securely, they must also be transmitted securely to CBO. For some data, such as tax data collected by the IRS and shared with CBO through SSA, an agreement details how to transmit the data using a secure server in addition to the agreement about data use.

Restricted data are stored in different ways at CBO, as agreed to in particular data use agreements. Some data are stored in separate servers after being transmitted

to CBO. Other data are stored in physical hard drives, either external hard drives or on a CBO computer.

Once the data are accessible from CBO's IT infrastructure, either physically at the Ford House Office Building or remotely, many agreements specify security protocols that must be implemented. Many protocols relate to physical access to the data. Examples include requiring locks on offices, electronic credential control for offices or office suites, and privacy shields on monitors. Sometimes hard copies of the data must be stored in a safe, or a separate personal computer has to be used solely for the purpose of accessing a data set, and it is not allowed to be connected to the Internet. Other requirements include additional log-in credentials, required training for analysts who use particular data, and additional security clearances or background checks.

Accessing Restricted Data Through the Data Owner

Some data are accessed through the IT infrastructure provided by the organization that owns the data. In the past, CBO's employees have been treated in the same manner as contractors by other agencies, such as the Federal Emergency Management Agency, the Department of Defense, or the Census Bureau. Specifically, CBO's employees have been able to log into those agencies' IT systems and access data through their infrastructure. Although it is possible that such arrangements may change in the future, the Census Bureau has built a Federal Statistical Research Data Center network, and CBO expects that it will be able to access restricted census data through that network for the foreseeable future.

Accessing data through the data owners' systems has some benefits. If data are updated, those changes are incorporated without any additional work from CBO's analysts. It also eases the ability of analysts to request additional data if they become necessary. It can facilitate communication between technical analysts at CBO and other agencies, allowing faster responses to questions about the data or documentation. It also means that CBO does not have to maintain the security of those data, which lowers costs for the agency.

There are also some drawbacks to accessing data through another agency's infrastructure. If the agency requires CBO's employees to access the data at its physical location, that can be cumbersome and inhibit CBO's ability

to be responsive to the Congress. Many times, an agency requires CBO's analysts to obtain badges or controlled access cards. Learning another agency's IT infrastructure and learning whom to contact for specific questions can be difficult, especially when someone seeks to use the restricted data for the first time. Furthermore, because the data are often collected for purposes of program administration rather than to facilitate statistical analysis, they may not be organized in a way that is useful for that purpose or may not be well-documented. Work may be required to enhance usability, which increases the time needed for analysis.

An additional concern about accessing data through another agency's infrastructure is that CBO also has an obligation to protect the confidential information of the Congress. Most often that information relates to the scope of legislative proposals that have not been made public. If CBO accesses data through another agency's infrastructure, it needs to ensure that the agency cannot glean the details of a Congressional proposal from the analysis.

Accessing Restricted Data Through a Third Party

Some data are accessed through a third party, such as a contractor of the agency that owns the data. For example, Acumen maintains restricted CMS data for Medicare and Medicaid claims, which it then securely transfers to CBO. Some possible projects would require CBO to access data directly through a secure enclave hosted by Acumen. Other examples of third parties that provide access to public data are IPUMS and ICPSR.

Benefits and drawbacks to accessing data through a third party are similar to those associated with accessing the data through the data owners' systems. However, with a third party, the data may be easier to use because they are typically well-documented and may have been processed to a certain extent for research purposes (because the third party is usually hosting the data for many different users). The data are also typically accessible remotely. However, accessing data through a third party may require additional resources—in dollar terms (to pay fees) or in terms of time (to create data use agreements).

Challenges and Opportunities During the Coronavirus Pandemic

Many security protocols related to data revolve around the physical security of the space where the data are

accessed and the physical storage of the data. Those issues created challenges when CBO's workforce began working from home in March 2020 in response to the 2020–2021 coronavirus pandemic. The agency has addressed issues related to data access in several ways:

- Arranging schedules so that employees visit the Ford House Office Building at different times to allow for social distancing;
- Shifting work to projects that use publicly available data and data that can be accessed remotely; and
- Updating agreements to allow analysts to work with restricted data remotely.

Moving forward, having agreements in place that allow CBO's analysts remote access to restricted data will increase the agency's responsiveness to the Congress. However, it also means that the agency will need to continue to invest in certain physical security measures, such as locking doors or providing screen protectors, both in the Ford House Office Building and in off-site work areas.

A particular opportunity to better access restricted data from the Census Bureau arose during the pandemic. The agreement in place before the pandemic required CBO's analysts to travel to the Census Bureau's headquarters in Suitland, Maryland, to access restricted Census and commingled data. However, the bureau was able to create temporary agreements allowing some researchers, including some of CBO's analysts, to log into the Census Bureau infrastructure through their Virtual Desktop infrastructure. The greater ease accessing the restricted microdata has facilitated progress on projects that use those data, including commingled survey and tax data.

Another opportunity to access tax data remotely arose during the pandemic. The data that CBO receives from SSA under title 26 were moved from an isolated system on-site to an approved system in the cloud, allowing CBO's analysts to access the data remotely. In the pandemic environment, that allows staff to be more responsive when there are limits to working in the physical office space. Other agencies, including the IRS and JCT, have been particularly helpful in ensuring CBO's access to data while keeping employees safe.

Confidentiality of Data

Federal agencies collect data for many reasons. Whether collected for administrative or statistical purposes, those data often contain sensitive information. That information could include names, birth dates, Social Security numbers, employer identification numbers, sensitive financial or medical information, trade secrets, proprietary business information, or classified information. Protecting the confidential information of people and institutions that give such data to federal agencies is key for maintaining the public's trust and for agencies to fulfill their missions.

Protecting Confidentiality

Data collected for statistical uses are protected explicitly by law, and CBO is obligated to protect data in the same way that other federal agencies do.²¹ The Confidential Information Protection and Statistical Efficiency Act of 2002 requires that when data are acquired by a statistical agency (or unit) under a pledge of confidentiality and for statistical purposes, the data may be used only for those purposes.²² It also requires the agencies to protect the confidentiality of personally identifiable information acquired for statistical purposes by adhering to principles designed to safeguard such information.²³ Laws such as title 13 and title 26 require strict protection and restrict publication of statistics that could lead to reidentification of individuals or firms.²⁴ They include strict penalties for unlawful disclosure of such information.

CBO protects the confidentiality of data in the same way as the agencies that produce them. To facilitate that, the agencies' technical experts have to explain the guidelines protecting their data. Sometimes they also review CBO's

completed work to ensure that legal protections have been properly applied.

To ensure that processes have been properly implemented and that data are not inadvertently disclosed, some agencies have set up a formal review process. Examples include the Department of Education and the Census Bureau. Estimates from those agencies cannot be published or used in models until after the review process is complete. Although this decreases the probability of inadvertent disclosure by CBO, it can make releasing information to the Congress less timely.

Differential Privacy

Certain agencies have been working to modernize their disclosure-avoidance practices. That is primarily because computing power has increased and because commercially available information might be linked with statistical information to identify individuals.²⁵

The Census Bureau has begun implementing a method known as differential privacy. Differential privacy injects statistical "noise" into the tabulated data, limiting the probability that any individual could be identified in the data set. Microdata are then created to be consistent with reported tabulations.²⁶ A greater amount of noise provides more protection against disclosure but less accuracy of the resulting data. The bureau plans to use that method for products associated with the 2020 decennial census and will adapt it for surveys in the future. The level of noise for the forthcoming redistricting data (data used to redraw legislative districts every 10 years) has been determined by the Census Bureau's Data

21. Sec. 203(e) of the Congressional Budget Act of 1974, P.L. 93-344 (codified at 2 U.S.C. §603(e) (2018)).

22. Sec. 512 of the Confidential Information Protection and Statistical Efficiency Act of 2002, P.L. 107-347 (codified at 44 U.S.C. §3501 (2018)).

23. Sec. 523 of the Confidential Information Protection and Statistical Efficiency Act of 2002, P.L. 107-347 (codified at 44 U.S.C. §3501 note (2018)). See Federal Committee on Statistical Methodology, *Report on Statistical Disclosure Limitation Methodology*, Statistical Policy Working Paper 22, second version (Office of Management and Budget, December 2005), https://nces.ed.gov/FCSM/pdf/SPWP22_rev.pdf (746 KB).

24. For examples, see 13 U.S.C. §214 (2018) and 26 U.S.C. §7213 (2018).

25. For examples, see Michael B. Hawes, "Implementing Differential Privacy: Seven Lessons From the 2020 United States Census," *Harvard Data Science Review*, Issue 2.2 (April 2020), <https://doi.org/10.1162/99608f92.353c6f99>; and John M. Abowd and Victoria A. Velkoff, "Modernizing Disclosure Avoidance: What We've Learned, Where We Are Now," *Census Blogs: Research Matters* (blog entry, March 13, 2020), <https://tinyurl.com/38wubv9r>.

26. Further explanation is available from Philip Leclerc, Mathematical Statistician, Center for Enterprise Dissemination—Disclosure Avoidance, "Generating Microdata With Complex Invariants Under Differential Privacy" (presentation to the 2019 Joint Statistical Meetings, July 2019), <https://tinyurl.com/jy2wbep4>.

Stewardship Executive Committee, which sets the formal “privacy budget.”²⁷

Broader implementation of those and other similar methods will negatively affect the quality of the publicly available data for CBO’s analyses. Implementation of enhanced privacy protection will also increase the time needed for some analysis. The time required might increase because agencies would validate the analysis using their own internal data or because CBO would need to request use of the internal data directly to increase accuracy. However, that method of privacy protection for public data helps CBO’s analysts ensure that they protect the data at the same level of confidentiality as the agencies that produce the data. CBO can also assist agencies as they implement differential privacy. When agencies test their disclosure-avoidance processes to ensure that key relationships from the underlying data are maintained, CBO can provide feedback about the relationships that are important for the work of the Congress.

27. See Testimony of Wilbur L. Ross Jr., Secretary of Commerce, before the House Committee on Oversight and Reform (March 14, 2019), <https://go.usa.gov/xHF8G>; and Declaration of John M. Abowd, Chief Scientist and Associate Director for Research and Methodology at the U.S. Census Bureau, *State of Alabama v. United States Department of Commerce*, U.S. District Court for the Middle District of Alabama, Eastern Division (April 13, 2021), p. 71, <https://tinyurl.com/exzkt6s9>.

This Congressional Budget Office report was prepared to inform the Congress. In keeping with CBO’s mandate to provide objective, impartial analysis, the report contains no recommendations.

Rebecca Heller and Mark Hadley prepared the report, with guidance from Xiaotong Niu and Julie Topoleski. Useful comments were provided by Alissa Ardito Ashcroft (formerly of CBO), Joseph E. Evans Jr., Ann E. Futrell, Tamara Hayford, Joseph Kile, Deborah Kilroe, Kevin Laden, David Rafferty, Davis Riley, Chayim Rosito, and Rebecca Verreau of CBO, and Thomas A. Barthold of the Joint Committee on Taxation. (The assistance of an external reviewer implies no responsibility for the final product, which rests solely with CBO.) Julia Heinzl fact-checked the report.

Jeffrey Kling and Robert Sunshine reviewed the report. Loretta Lettner was the editor, and Jorge Salazar was the graphics editor. This document is available at www.cbo.gov/publication/57150.

CBO continually seeks feedback to make its work as useful as possible. Please send any comments to communications@cbo.gov.



Phillip L. Swagel
Director

