
# Researcher Access to Restricted Government Data

Jen Rosiere Reynolds, Aditi Bawa, and Kamya Yadav
Carnegie Endowment for International Peace

June 2, 2022

**Speeding Research on the Information Environment**
**researchaccelerator.org**

## Abstract

We analyzed 31 procedures researchers use to access restricted data from selected government institutions in five countries.[1] Various organizations adjudicated data access within our sample, including departments of labor, health, census, and economics within Australia, Canada, New Zealand, the United Kingdom, and the United States. This paper defines restricted data as data that cannot be released directly to public researchers due to possible risks to study participants and the confidentiality promised to them.[2] Researcher eligibility requirements ranged from zero to numerous conditions and included either individual researcher or project requirements. Every one of the 31 procedures we studied used one or more protection mechanisms to provide oversight and guard data from malicious or negligent misuse. These structural, process, or programmatic approaches were often part of a multi-layered approach. All 31 of these processes required approval, and the majority required a binding agreement, pre-publication review, and an advisor before or during to guide the process. Further frequent commonalities among the examined procedures included an ethics approval, a background check, and a researcher fee. Finally, we offer a possible model for the Institute for Research on the Information Environment based on all access models studied.


[1] It is important to note that these 31 analyzed procedures do not constitute all types of procedures researchers use to access restricted data, but represent *options* for data access procedures that the Institute for Research on the Information Environment can adopt.

[2] "Access to Restricted Data". *National Institutes of Health (NIH)National Institute on Aging (NIA).* https://www.nia.nih.gov/research/dbsr/access-restricted-data


**Funding:** This research was generously supported by Microsoft, Craig Newmark Philanthropies, the John S. and James L. Knight Foundation, and the William and Flora Hewlett Foundation. This work was initiated as part of a joint project between the Partnership for Countering Influence Operations at the Carnegie Endowment for International Peace and the Empirical Studies of Conflict Project at Princeton University. **Author contributions:** JRR drafted the manuscript and provided oversight and leadership responsibility for the research activity planning and execution. AB collected data and wrote original draft. KY curated data. **Data and materials availability:** All data required to evaluate the conclusions are present in the paper or supplemental materials. Additional data related to this paper may be requested from the authors.

**Executive Summary**

1. We studied 31 restricted data access procedures across five governments: Australia, Canada, New Zealand, the United Kingdom, and the United States. Processes often covered more than one restricted dataset. These data were related to various topics: labor, demographics, judicial systems, health, housing, and education.
2. The access models we examined had various eligibility requirements for researchers. The most common prerequisites were mission or value alignment, educational or skill, and organizational affiliation.
3. We found a range of protection mechanisms aimed to provide oversight and protect data from malicious or negligent misuse and categorized them as "structural," "process," or "programmatic." Government entities that adjudicate restricted data access often used more than one of these mechanisms.
4. The majority mandated a binding agreement and output review, and most provided or required an advisor before or during the process. Further frequent commonalities included ethics approval, a background check, and a fee.
5. Based on the access models studied, we outline a possible model for the Institute for Research on the Information Environment:
    a. A researcher submits a research proposal that includes an ethics committee approval. This ethics committee may be unique to IRIE or part of a partner institution, such as a university.
    b. The proposal undergoes a committee review (possibly just meaning more than one IRIE staff member reviews). It is accepted, rejected, or researchers are requested to provide clarification or revise and resubmit.
    c. Once the project has all necessary approvals, the researchers undergo a background check to verify their identity, educational and professional credentials, and personal and professional references.
    d. Researchers must then attend an orientation session, during which they sign a contract.
    e. All researchers must complete annual training in proper data stewardship.
    f. When the researcher has finished their work, an IRIE advisor will conduct a disclosure or confidentiality review of all project outputs to protect data confidentiality. This advisor could be a faculty affiliate or IRIE staff.

## Introduction

To assess possible access models for the Institute for Research on the Information Environment (IRIE), we collected information on how 20 government bureaucratic agencies in English-speaking democracies manage access to restricted government data. There are various components of vetting researchers and allowing them access to sensitive information collected about individuals. All publicly available applications are listed and linked in Appendix A.2. Our research identified various approaches to eligibility requirements, protection mechanisms, and access with commonalities of a multi-step undertaking, similar prerequisites and a multi-layered approach to protect privacy, provide oversight, and prevent misuse.

Of note, informing these nations' processes, New Zealand and the United Kingdom had overarching data privacy acts specifying the legal basis for accessing restricted data across agencies. Australia,[3] Canada,[4] and New Zealand[5] had designated data and statistics agencies that set regulations for researchers accessing data.

## Methodology

Using our team's knowledge of sensitive government data sets and internet searches, we 20 identified high-functioning government institutions in English-speaking democracies that seek to collect and make available sensitive information tied to the identity of individuals or firms. Collecting and sharing restricted government data is analogous to doing so with social media data, as many of the same privacy, legal and ethical issues arise. We researched agencies with authority over restricted data within five governments (Australia, Canada, New Zealand, the United Kingdom, and the United States), used open internet research to examine the researcher access procedures, and identified 31 unique processes. In cases where the approach was unclear, we contacted the organization via email.

We sought to answer the following questions:

1. What is the application process and criteria for access?
2. What kinds of sensitive data do they give access to?
3. How are projects using the data funded?
4. How do they manage security (NDAs, physical separation such as Census Secure Data Centers, or other requirements)?

The list of included restricted data sets (see Appendix A.1) is not exhaustive, as we could not find publicly available data on all variables.

---

[3] "DataLab - Topics". *Australian Bureau of Statistics (ABS)* (11 August 2021). https://www.abs.gov.au/statistics/microdata-tablebuilder/datalab/topics

[4]"Statistics Canada". *Statistics Canada* (16 June 2022). https://www.statcan.gc.ca/en/start

[5] "Statistics". *New Zealand Government Stats NZ* (March 2022). https://www.stats.govt.nz/

## Eligibility Requirements

Across the 31 restricted data access procedures we studied, most had eligibility requirements for either individual researchers or projects. Detailed project eligibility requirements were common but not universal.

1. 10 processes called for some alignment with the data-owning agency or entity, including four processes that required the project be aligned with their mission or benefit their programs. Six mandated the project, or researcher, to be generally working for the public good.
2. Three processes required a technical review to demonstrate the scientific or technical merit of a proposed project or how it would advance scientific knowledge. Two processes explicitly required that the project be feasible to be approved.

Requirements regarding the researcher also ranged in detail.

1. Six processes had specific educational or experience prerequisites, such as statistical coursework. Three similarly mentioned a requirement to have the skills needed to work with the data but did not specify educational or training constraints.
2. Five procedures included organizational requirements, mandating that researchers work under the direction or supervision of, or be affiliated with, a public, private, or academic organization.

Requiring skills instead of coursework is more inclusive toward researchers with untraditional backgrounds but may be more challenging to vet. An affiliation constraint, as described above, excludes independent researchers if the granting organization does not allow exceptions. However, an organizational connection indicates that the individual accessing the restricted data has been vetted by or is tied to an institution's reputation.

## Protection Mechanisms

The studied government organizations that adjudicated restricted data access employed a range of protection mechanisms and often used a multi-layered approach. These elements aimed to provide oversight and protect data from malicious or negligent misuse. These approaches can be broken into three categories–structural, process, or programmatic–and every one of the 31 approaches we studied used one or more.

Structural protection mechanisms control the manner through which researchers access data. The most common structural tool is the use of access points. This mechanism requires researchers to access data through a secure virtual or physical site, thus enabling review, access control, and auditing. 26 of the 31 processes used a structural protection mechanism.

For example, United States Federal Statistical Research Data Centers (FSRDCs) are facilities housed in over 50 partner institutions (including universities and non-profit research institutions) that meet particular security requirements for access to restricted US data. FSRDCs are equipped with software that enables researchers to analyze large datasets in a secure environment. All of the agencies analyzed within the United States allowed data access within an FSRDC. These are similar to Research Data Centers (RDCs) in Canada, university-based laboratories that offer researchers access to confidential microdata, fully-equipped workstations, statistical software, and technical support. An alternative to this physical access model is the virtual model. For example, the Australian Bureau of Statistics DataLab is an online platform where all activity is recorded, reviewed, and subject to auditing. If an audit showed that a user has breached the DataLab conditions of use, the user's access could be withdrawn. Similarly, the SecureLab affiliated with the UK Data Service, UK's largest qualitative and quantitative social science data repository, is a remote access lab withphysical locations across the UK so that researchers can choose their preferred access mode.

Another structural protection mechanism is the removal of all identifying information within the dataset before researchers gain access. This design employs the highest level of security, as it removes any possibility of researchers misusing data, but it severely limits possible analyses. Certain research questions require identifying information for effective analysis. For example, a person may use the same username on two different social media platforms; if a researcher is trying to see whether an observed message reflects a trend or just a single individual, knowing the identifying username is integral.

Process mechanisms are actions or steps a researcher must take to access restricted data. In our study, these mechanisms included research plan review and approval as well as disclosure plan approval. 22 of the 31 access procedures used one or more process protection mechanisms.

Programmatic protection mechanisms were the least common type of requirement. These include the following:

1. The non-retention or destruction of data after a period of time (such as one year or upon finishing research);
2. A condition that research must align with the granting agency's mission; and
3. A general requirement that research must be for the public good.

Only seven processes used a programmatic mechanism, and all were coupled with either structural or structural and process mechanisms.

**Access Process Features**

All 31 of these processes required approval, and many centralized their list of requirements in one application requesting authorization. 20 of the 31 access procedures required a signed agreement or oath. 19 required a confidentiality or disclosure review before a researcher publicized any data or

analysis. Importantly, none of the processes we examined required a review on the nature of the results. 18 access procedures incorporated an analyst, liaison, or administrator before or during the application. 15 processes required ethics approval for data access. 12 required some type of background or security check, including the US Census Bureau's Special Sworn Status (SSS). Seven access procedures included charging a fee to researchers, while one process may charge for customized statistics and data. Finally, 11 processes required specific training to access the data.

*Binding Agreement*

20 of the 31 access procedures required a signed agreement or oath. Publicly available sample legal agreements are listed in Appendix A.1. 10 of these 20 signed agreements were a contract to comply with the terms and conditions of the research. The contracts could be between the granting agency and either the researcher or the researcher's sponsoring organization. These agreements could specify fees, time limits on data sharing, and ethics approvals.

The other 10 binding agreements were oaths or confidentiality agreements to protect the data, including agreements by the US Census bureau and RDC. Researchers must obtain SSS to access US census data. SSS individuals are sworn to protect the data for life, just as Census Bureau employees are, and are subject to the same legal obligations and penalties for misuse. Similarly, RDC required researchers to take an Oath of Office and Secrecy, requiring that they are held accountable for maintaining the data's confidentiality for life. If the researcher violates this oath, they lose their Statistics Canada privileges. Once Statistics Canada approves a research project, they create a contract between themselves and the researcher, like the signed agreements detailed above, to ensure compliance with terms and conditions.

Requiring researchers to explicitly commit to protecting data confidentiality may be an effective way to protect the individuals behind the data.

*Output Review*

Before a researcher could make any data or analysis public, 19 processes required a confidentiality or disclosure review, but none included a review on the nature of the results. 15 of the 19 output reviews pertained strictly to the confidentiality of study subjects. Of note in the remaining four procedures, the US National Center for Health Statistics reviewed all outputs for privacy as well as for alignment between the research questions and the work described in the approved proposal, and the UK's Ministry of Justice reviewed that no output jeopardizes the long-run value, integrity, and sustainability of the studies. Stats New Zealand published a detailed guide for researchers to self-regulate confidentiality, privacy, and security before requesting publication review. See Appendix A.3 for details.

*Access Support*

Eighteen access procedures incorporated an analyst, liaison, or administrator before or during the application process. An appointee serving as an intermediary between the researcher and the agency

who provides counsel may facilitate applications and reduce the administrative burden on both an unfamiliar researcher and an organization receiving unsuitable requests.
Various organizations used this model, including:

1. The United States' National Center for Health Statistics[6] required researchers to work with an analyst who facilitates proposal review, creates the dataset, conducts a disclosure review of the researcher's final output, and answers questions throughout the process.
2. RDCs employed analysts to advise the researcher on an application's feasibility and appropriateness. Additionally, these hanalysts controlled the output publicly released or published to ensure confidentiality.[7]
3. The Australian Department of Health[8] appointed delegates to work with applicants to identify and allow access to only the most essential and relevant data. These representatives took legal responsibility for the data released.

*Ethics Review*
15 processes required ethics approval for data access, with all but two requiring evidence of approval from an established ethics board. The two exceptions here are the:

1. Growing up in New Zealand Dataset Accredited Researcher procedure, which requested that a researcher gain ethics approval but did not require evidence of such, and
2. US Department of Health and Human Services' HIPAA Privacy Rule, whose process requested that the researcher provide either IRB approval, a written or oral disclosure that the data access was sought as a type of preparatory research, disclosure that the researcher was only seeking data of decedents, disclosure that the data set was limited, or that the data was about the researcher themselves.

*Background Checks*
12 processes required some background or security check, including special sworn status. These 12 procedures included the US Census Bureau, RDC, three others from the United States, two each from Australia, New Zealand and the UK, and one from Canada.

Special Sworn Status applicants had to undergo a medium-level FBI background check by submitting a

[6] "Federal Statistic RDC". *Centers for Disease Control and Prevention (CDC)* (6 May 2021). https://www.cdc.gov/rdc/b2accessmod/ACs220.htm
[7] "Guide for Researchers under Agreement with Statistics Canada". *Statistics Canada* (October 2005). https://cdn.dal.ca/content/dam/dalhousie/pdf/faculty/ardc/researcher-rechercheur-guide-eng.pdf
[8] "Data Access and Release Policy". *Australian Government Department of Health and Aged Care* (16 February 2018). https://www1.health.gov.au/internet/main/publishing.nsf/Content/Data-Access-Release-Policy

work and residence history and fingerprints.[9][10] RDC's background check considered the previous five years, including verification of identity and background, educational and professional credentials, personal and professional references, credit, and criminal records.[11]

*Fee for Access*

Seven access procedures included charging a fee to researchers, and one process may charge per extract or for requests for customized statistics and data. Of those seven processes with fees, three charged approximately between $4,000 and $11,000, two procedures charged under $1,000, and two did not specify an estimated cost but noted that costs must be reasonable per the New Zealand Official Information Act. Two of the seven charged based on the size of the data requested. Six of the seven itemized their fees; these lists included administrative support (confidentiality checking, amendments, renewals, special security status), onsite usage, technical assistance, and storage.

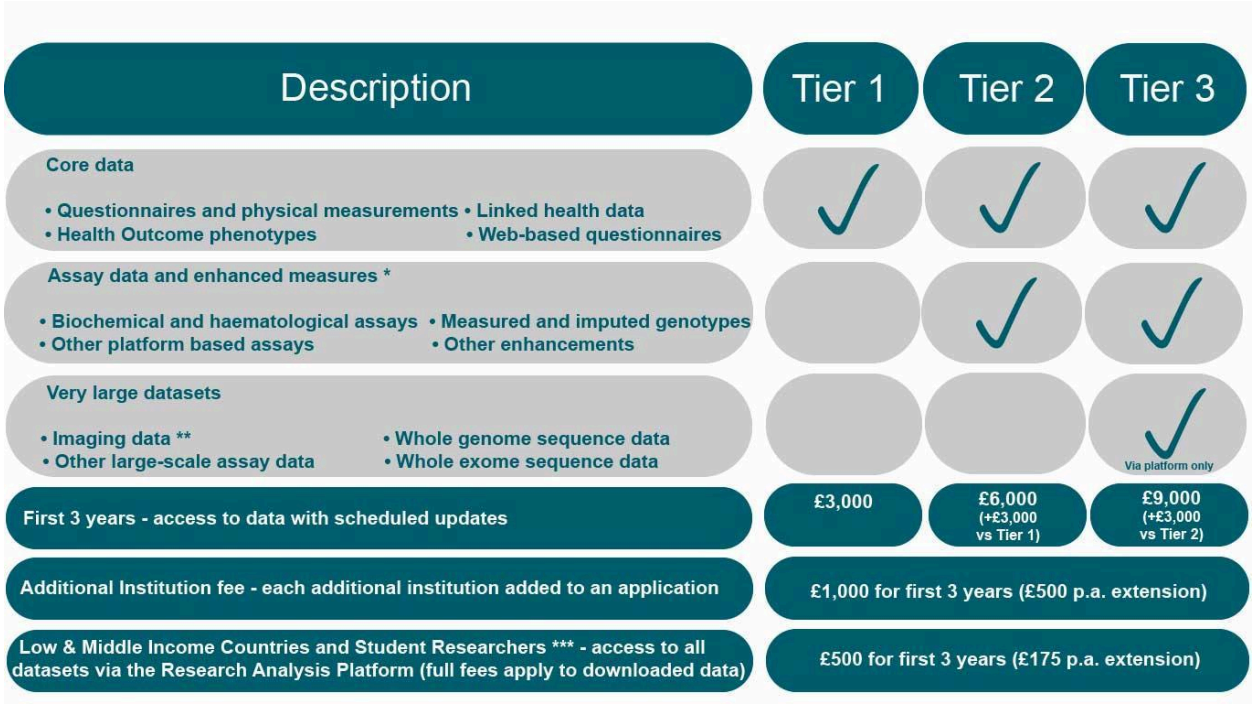
The UK Biobank had a unique fee structure, as outlined in Table 1, with costs paralleling the complexity of the data set and exceptions for students and researchers from low- and middle-income countries.

[9] Dudenhefer, Paul. "Accessing Important Census Data, Confidentially". *Duke Today* (30 May 2013). https://today.duke.edu/2013/05/confidential

[10] " Guidelines for the Development and Operation of a Federal Statistical Research Data Center (RDC) or RDC Branch". *US Census Bureau* (18 March 2016). https://www.census.gov/content/dam/Census/about/about-the-bureau/fsrdc/Guidelines_for_RDC_Development_and_Operations_FINAL_2016.pdf

[11] "Standard on Security Screening". *Government of Canada Treasury Board of Canada* (20 October 2014). https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=28115#appB

**Table 1: Example UK Biobank Fee Structure**




**Access Process Examples**

*Statistics Canada*

The RDC Program provided researchers access to microdata from population and household surveys. The access process is as follows and featured approval, an advisor, a background check, special status, training, a confidentiality/binding agreement, an output review, and affiliation requirements.

1. All researchers had to submit a research proposal to an adjudicating committee.
2. The proposal was reviewed and accepted or rejected, or researchers were requested to provide clarification or revise and resubmit the proposal. Unfortunately, we could not identify publicly available details on who reviewed the proposals.
3. When the proposal was approved, researchers had to get security clearance from Statistics Canada (an enhanced reliability check with security clearance forms).
4. Researchers were then required to attend an orientation session, during which they signed their contract with Statistics Canada and took the Oath of Office and Secrecy.
5. After signing the Oath, researchers were deemed employees of Statistics Canada for the duration of their contract.
6. The researcher was required to deliver an end product at the end of the contract.

*National Health Services*

National Health Services from the UK adjudicated the use of many health-related datasets, including those related to COVID-19 (Hospitalization in English Surveillance System, Non-hospital Antigen

Testing Results, and Vaccination Status Data); the Adult Psychiatric Morbidity Survey; Cancer registration data; and Demographic data. The process outlined below included the features of approval, deidentification, an advisor, a confidentiality/binding agreement, an output review, and an ethics approval.

1. Researchers filled out an application that set out the nature of the requested data and the purpose for which it is being requested.
2. A case officer reviewed the application and, if necessary, scheduled an appointment to discuss any remaining tasks that must be completed for the application to progress.
3. The NHS signed the Data Sharing Agreement.
4. The researcher reviewed the agreement, provided a purchase order number, and signed.
5. The data, upheld as appropriate in the case of any patient objections, was produced, reviewed, and signed off on by NHS Digital or the data service access was granted. The data was made available either by Secure File Transfer or through the NHS' Data Access Environment.

**Possible Model**

One proposed access model follows. It does not include a fee structure, because this paper does not compare financial models. However, it does include the common features of a binding agreement, output review, administrative support during the process, ethics review, a background check, and training.

1. A researcher submits a research proposal that includes an ethics committee approval. This ethics committee may be unique to IRIE or may be part of a partner institution, such as a university.
2. The proposal undergoes a committee review (possibly just meaning more than one IRIE staff member reviews). It is accepted, rejected, or researchers are requested to provide clarification or to revise and resubmit.
3. Once the project has all necessary approvals, the researchers undergo a background check to verify identity, educational and professional credentials, and personal and professional references.
4. Researchers must then attend an orientation session, during which they sign a contract.
5. All researchers must complete annual training in proper data stewardship.
6. When the researcher has finished their work, an IRIE advisor will conduct a disclosure or confidentiality review of all project outputs to protect data confidentiality. This advisor could be a faculty affiliate or IRIE staff.

While this model provides a start, it needs further refinement and yields various questions. For example, who would comprise the committee to review and approve research proposals? Governments that adjudicate the researched models have different capacities to execute background checks. Who or what would evaluate background checks if IRIE uses these as a protection mechanism? Who administers data stewardship training?

## Conclusion

Each of the 31 processes examined required approval, and the majority required a binding agreement, output review, and an advisor before or during the undertaking. An appointee, serving as an intermediary between the researcher and the agency, may provide counsel to facilitate applications and reduce the administrative burden on both researchers who are unfamiliar with the process and an organization receiving unsuitable requests. Further frequent commonalities in access procedures included ethics approval, a background check, and a researcher fee.

Among the examined access models, the number of eligibility requirements ranged from zero to numerous. At a minimum, requiring the specific skills needed to work with the data seems both effective in preventing difficulties after beginning work and inclusive toward researchers with untraditional backgrounds.

Further, all 31 processes employed one or more protection mechanisms, often part of a multi-layered approach. This paper does not recommend in-person or virtual access models, as making such a recommendation would be premature before IRIE specifies its products and customer needs. However, we do recommend process protection mechanisms.

As a result of this research, a few questions emerge that a researcher may be able to answer through qualitative interviews: What are the missing efficiencies within these procedures? For example, are any steps more or less helpful in vetting access? Could or should an ideal process be more collaborative?

**Appendix**

*A.1 Included Datasets*

| **Agency/Process** | **Country** | **Datasets** |
|---|---|---|
| AURIN | Australia | Restricted Access geoscape data including buildings, trees, and surface cover |
| AURIN | Australia | Restricted Access APM Point Level Data (the enriched point location for properties advertised for rent, for sale, and sold in Australia). |
| AURIN | Australia | Restricted Access Landgate Data |
| Australian Bureau of Statistics (ABS) - DataLab | Australia | Census data, health, education, labor force, Aboriginal and Torres Strait Islander people data, migrants, crime, business, disabilities, aging, and carers data |
| Department of Health | Australia | Health program and health performance data |
| Independent Hospital Pricing Authority (IHPA) | Australia | Hospital cost data<br>Admitted patient activity<br>Emergency patient activity<br>Non-admitted patient activity<br>Mental health patient activity<br>Teaching, training and research activity |
| Canadian Health Measures Survey (CHMS) Biobank | Canada | Specimens of plasma, whole blood, DNA, urine, and serum |
| Statistics Canada - Research Data Centers (RDC) Program | Canada | Microdata from population and household surveys |
| Growing up in New Zealand Dataset - Accredited Researcher | New Zealand | Longitudinal study of child development, including data on health/wellbeing, psychosocial development, education, family, culture/identity, societal context, and neighborhood environment |
| Growing up in New Zealand Dataset - External Researcher | New Zealand | Longitudinal study of child development, including data on health/wellbeing, psychosocial development, education, family, culture/identity, societal context, and neighborhood environment |

| | | |
|---|---|---|
| Growing up in New Zealand Dataset - GUiNZ member | New Zealand | Longitudinal study of child development, including data on health/wellbeing, psychosocial development, education, family, culture/identity, societal context, and neighborhood environment |
| Ministry for Children - Oranga Tamariki | New Zealand | Data related to children in care, harm and abuse, health, parenting, and youth justice |
| Ministry of Health - Deidentified Data | New Zealand | Datasets include cancer, laboratory testing warehouse, mortality, National Booking Reporting System, National Health Index, and Medical Warnings System, National Immunisation Register, National Non-Admitted Patient Collection |
| Ministry of Health - Identifiable Data | New Zealand | Data can be requested and accessed from the datasets below: cancer, laboratory testing warehouse, mortality, National Booking Reporting System, National Health Index and Medical Warnings System, National Immunisation Register, National Non-Admitted Patient Collection. |
| Stats NZ - New Zealand's Official Data Agency | New Zealand | Integrated Data Infrastructure (IDI) and Stats NZ Surveys, include housing, people/communities, education/training, income/work, benefits/social services, population, health, and justice data |
| Ministry of Justice | United Kingdom | Her Majesty's Courts & Tribunals Service Data (HMCTS) Data is on the operation of the criminal and civil justice systems |
| Ministry of Justice | United Kingdom | Department for Education (MoJ-DfE datashare only) |
| Ministry of Justice | United Kingdom | Her Majesty's Prison and Probation Service (Data First Datasets only, MoJ Data First Criminal Courts, Prisons and Probation Linked Dataset, the MoJ Data First Prisoner Custodial Journey Level Dataset, and the MoJ Data First Probation Dataset) (HMPPS) |

| | | |
|---|---|---|
| National Health Services | United Kingdom | Datasets include: Coronavirus datasets (Hospitalization in English Surveillance System, Non-hospital Antigen Testing Results, Vaccination Status Data); Adult Psychiatric Morbidity Survey; Cancer registration data; Demographic data |
| UK Biobank | United Kingdom | Genetic data, health records, imaging data, health linkages, biomarkers, activity monitors |
| UK Data Service | United Kingdom | Large surveys funded by the UK Economic and Social Research Council, such as longitudinal cohort studies from the Centre for Longitudinal Studies – for example, the Millennium Cohort Study and Next Steps – and the UK's largest longitudinal household survey, Understanding Society. |
| Agency for Healthcare Research and Quality (AHRQ) | United States | Restricted Data from the Medical Expenditure Panel Survey, a set of surveys of and about families and individuals, their medical providers, and employers across the United States. |
| Bureau of Economic Analysis | United States | Foreign direct investment, the activities of multinational enterprises, and international trade in services |
| Bureau of Labor Statistics - FSRDC Access | United States | Census of Fatal Occupational Injuries<br>International Price Program<br>National Compensation Survey<br>National Longitudinal Surveys (NLS) Original Cohorts: Mature and Young Women, and Older and Young Men files with Geographic Variables<br>National Longitudinal Surveys of Youth 1979 (NLSY79) Zip Code and Census Tract Files<br>National Longitudinal Surveys of Youth 1997 (NLSY97) Zip Code and Census Tract Files<br>NLSY97 School Surveys<br>Occupational Requirements Survey (ORS)<br>Producer Price Indexes (PPI)<br>Survey of Occupational Injuries and Illnesses (SOII) |
| Bureau of Labor Statistics - Offsite Limited Access | United States | Census of Fatal Occupational Injuries Research File |

| | | |
|---|---|---|
| Bureau of Labor Statistics (BLS) - Offsite Limited Access | United States | National Longitudinal Surveys of Youth Geocode Data (information at multiple points in time on the labor market activities and other significant life events with locational information) |
| Bureau of Labor Statistics (BLS) - Onsite Access | United States | Census of Fatal Occupational Injuries<br>CFOI Micro Fatality Research File<br>Consumer Expenditure Survey (CE)<br>Consumer Price Indexes (CPI)<br>Current Employment Statistics (CES): National and State and Area<br>International Price Program (IPP)<br>Job Openings and Labor Turnovers Survey (JOLTS)<br>Mass Layoff Statistics (MLS)<br>National Compensation Survey (NCS)<br>National Longitudinal Surveys (NLS) Original Cohorts: Mature and Young Women, and Older and Young Men files with Geographic Variables<br>National Longitudinal Surveys Geocode Data<br>National Longitudinal Surveys of Youth 1979 (NLSY79) Zip Code and Census Tract Files<br>National Longitudinal Surveys of Youth 1997 (NLSY97) Zip Code and Census Tract Files<br>NLSY97 School Surveys<br>Occupational Employment Statistics (OES)<br>Occupational Requirements Survey (ORS)<br>Producer Price Indexes (PPI)<br>Quarterly Census of Employment and Wages (QCEW) and Longitudinal Database of Establishments (LDE)<br>Surveys of Employer Provided Training (1993 and 1995)<br>Survey of Occupational Injuries and Illnesses (SOII) |
| HIPAA Privacy Rule - Health and Human Services | United States | Protected health information data |

| | | |
|---|---|---|
| National Center for Health Statistics | United States | Geographic Codes for all NCHS Surveys<br>National Health and Nutrition Examination Survey (NHANES)<br>National Health Care Surveys<br>National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS)<br>National Hospital Discharge Survey (NHDS)<br>National Nursing Home Survey (NNHS) and National Nursing Assistant Survey (NNAS)<br>National Home and Hospice Care Survey (NHHCS) and National Home Health Aide Survey (NHHAS)<br>National Survey of Residential Care Facilities (NSRCF)<br>National Study of Long-Term Care Providers (NSLTCP)<br>National Hospital Care Survey (NHCS)<br>National Health Interview Survey (NHIS)<br>National Survey of Family Growth (NSFG)<br>State and Local Area Integrated Telephone Survey (SLAITS)<br>National Survey of Children's Health (NSCH)<br>National Survey of Children with Special Health Care Needs (CSHCN)<br>NCHS Data Linkage Activities<br>Linked Mortality Data Products<br>Linked Medicare Enrollment and Claims Files Data<br>Linked Medicaid Enrollment and Claims Data<br>Linked Social Security Benefit History Data<br>Linked Housing Assistance Program Files Data<br>Linked USRDS ESRD Data<br>National Vital Statistics System (NVSS) Data Release and Access Policy<br>Birth Data Files<br>National Maternal and Infant Health Survey<br>Mortality Data Files<br>Multiple Cause of Death Data File<br>Drug Involved Mortality Restricted Variables<br>Redacted Death Certificate Literal Text File<br>Research and Development Survey (RANDS) |

| | | |
|---|---|---|
| National Center for Science and Engineering Statistics | United States | Early Career Doctorate Survey (ECDS) - This survey gathers information about individuals who earned their first doctoral degree (PhD, MD, or equivalent) in the past 10 years and work at academic institutions and federally funded research and development centers.<br>National Survey of College Graduates (NSCG) - This is a study by the National Science Foundation, and provides information about the education and career paths of the Nation's college graduates.<br>National Survey of Recent College Graduates (NSRCG) - This was a cross-sectional biennial (1973-2010) survey that provided demographic and career information about individuals holding a bachelor's or master's degree in a science, engineering, or health field from a U.S. academic institution.<br>Scientists and Engineers Statistical Data System (SESTAT) Integrated File - This survey is a source of longitudinal information on the education and employment of the college-educated U.S. science and engineering workforce.<br>Survey of Doctorate Recipients (SDR) - This survey provides demographic, education, and career history information from individuals with a U.S. research doctoral degree in a science, engineering, or health (SEH) field.<br>Survey of Earned Doctorates (SED)/Doctorate Record File (DRF) - This is an annual census conducted since 1957 of all individuals receiving a research doctorate from an accredited U.S. institution in a given academic year. |
| US Census Bureau | United States | Inventory List<br>Demographic Data<br>Economic Data<br>LEHD Data - Firms and workers sourced from administrative data<br>UMETRICS Data - Microdata on research grants and projects linked to Census Bureau data on |

| | | |
|---|---|---|
| | | employment and employers |

*A.2 Example Binding Agreements*

| Government | Agency | Legal Agreement |
|---|---|---|
| United Kingdom | UK Data Service | Secure Access User Agreement |
| United Kingdom | National Health Service | Data Sharing Agreement |
| United States | National Center for Health Statistics | Data Use Agreement |
| Canada | Statistics Canada | Oath of Office |

*A.3 Example Applications*

| Government | Agency | Application |
|---|---|---|
| Australia | Department of Health | Requesting Health Statistics and Data |
| Australia | Independent Hospital Pricing Authority (IHPA) | Research Data Request Form |
| Australia | Australian Bureau of Statistics (ABS) | Applying for and using DataLab |
| Canada | Statistics Canada - Research Data Centers (RDC) Program | Research Data Centres: Application Process |
| Canada | Canadian Health Measures Survey (CHMS) Biobank | Application Process |

| | | |
|---|---|---|
| New Zealand | Growing up in New Zealand Dataset | The Data Access Application Process |
| New Zealand | Stats NZ - New Zealand's Official Data Agency | Apply to Use Microdata for Research |
| New Zealand | Ministry for Children - Oranga Tamariki | Application Form |
| New Zealand | Ministry of Health | Data/Information Requests |
| United Kingdom | Ministry of Justice | Accessing Data via the Research Accreditation Framework |
| United Kingdom | UK GDPR | How do we Recognize and Subject Access Requests? |
| United Kingdom | UK Data Service | Application Guide |
| United Kingdom | UK Biobank | Material Transfer Agreement |
| United Kingdom | National Health Services | Data Access Request Service Process |
| United States | US Census Bureau | ERD Evidence-Building Project Proposal |
| United States | Bureau of Labor Statistics (BLS) | BLS Restricted Data Access Application |

| | | |
|---|---|---|
| United States | Agency for Healthcare Research and Quality (AHRQ) | AHRQ Data Center Application |
| United States | Bureau of Economic Analysis | How Do I Propose a Project? |
| United States | National Center for Health Statistics | Proposal Format |
| United States | National Center for Science and Engineering Statistics | Data Application |

*A.4 Example guide for confidentiality, privacy, and security before output review*
Microdata output guide

*A.5 Data*
Access to Data in Gov. Institutions