Tiered Access to Research Data for Secondary Analysis

John Marcotte, Sarah Rush & Kelly Ogden-Schuette
Data Sharing for Demographic Research
University of Michigan

As the richness of research data for secondary analysis has grown, disclosure risk has also increased. Research data have expanded in their gradation of the risks associated with both re-identification and harm, which has created a need for multiple levels of access controls beyond public and restricted access. Public-access data have typically been available for download from websites with minimal conditions on how data may be used while restricted-access data usually require an application and formal authorization process. The old paradigm of classifying data as public-access or restricted-access is no longer sufficient. Access to research data requires more nuance to ensure the protection of human subjects.

Throughout this paper, we characterize research data as information for producing aggregate results such as summary statistics and regression coefficients. These aggregate results must meet disclosure protection thresholds for cell sizes for tables, sample sizes for regressions, and other specified conditions such as the suppression of certain variables or disallowed sub-samples. Although research data may contain information about individuals and organizations, research data are not intended for identifying individuals or organizations.

In the last 20 years, several articles and reports develop frameworks for providing access to research data. In a 2002 report entitled *Restricted Access Procedures*[1], the Confidentiality and Data Access Committee identify Research Data Centers (RDC) as a primary method for providing access to restricted-access research data. The report also discusses remote access and online query systems as alternatives to RDC. At the time of this report, remote access systems were still in their infancy. Several years later, Kinney, Karr and Gonzales (2010)[2] discuss direct access through RDC and licensed access for researchers to analyze data on their own computers. Kinney, et al. also propose using tabular and synthetic data to mitigate disclosure risk. More recently, Desai, Ritchie, and Welpton (2016)[3] describe the Five Safes framework for data access. The framework, based on aspects related to the project, people, data, settings and output, can be a basis for designing tiered access but Desai, et al. do not define specific access levels.

[1] Federal Committee on Statistical Methodology: Confidentiality and Data Access Committee. (2002, April). *Restricted Access Procedures*. https://nces.ed.gov/FCSM/pdf/CDAC_RAP.pdf

[2] Kinney, S. K., Karr, A. F., & Gonzalez Jr., J. F. (2010). Data Confidentiality: The Next Five Years Summary and Guide to Papers. *Journal of Privacy and Confidentiality*, *1*(2). https://doi.org/10.29012/jpc.v1i2.569

[3] Desai, T., Ritchie, F., & Welpton, R. (n.d.). Five Safes: designing data access for research. In *Economics Working Paper Series 1601*. University of the West of England, Bristol. https://www2.uwe.ac.uk/faculties/bbs/Documents/1601.pdf

More than 1,800 public research data repositories are currently available to academia, government, and business.[4] Several of these repositories have established different levels of access that have some overlap with the seven tiers we propose in this paper. These levels usually focus on technology. While some repositories have offered tiered access, our unique contribution is how we define our tiers in terms of both human and technical controls to prevent the release of disclosive information.

Dataverse[5] offers tiered access but only specifies additional technical controls such as encryption and two-factor authentication. To our knowledge, Dataverse does not have a system in place to review output.

Guidance for Controllers on Data Security (February 2020)[6] from the **Irish Social Science Data Archive (ISSDA)[7]** is an excellent example of how repositories approach both physical security and the human factor. The Physical Security requirements overlap with many other repositories while the discussion of the human factor does not suggest controls beyond training, accountability, and continuity.

Slavkovic, Kinney and Karrin writing in *Chance* (2013)[8] cite the **National Opinion Research Center (NORC)[9]** Data Enclave as an example of an online data enclave. They describe both technical and human controls including how researchers access data over an encrypted connection and are unable to transfer any data, even via copy and paste, to their local computer. Moreover, output must be reviewed before release to researchers. They also discuss the cost of operations. Slavkovic, et al. also mention Census Research Data Centers, but they do not indicate what additional security controls that the Census Research Data Centers have over an online enclave.

According to Thissen and Mason in *Health Systems* (2019)[10], security controls for research data depend not only on the sensitivity of the information but also on regulations, requirements, or ethical constraints. Compliance with regulations is a key aspect of specifying controls. Thiessen and Mason do not discuss how compliance is ensured.

[4] Crosas, M. (n.d.). *CIO Review: Cloud Dataverse: A Data Repository Platform for the Cloud*. https://openstack.cioreview.com/cxoinsight/cloud-dataverse-a-data-repository-platform-for-the-cloud-nid-24199-cid-120.html

[5] *The Dataverse Project*. (n.d.). https://dataverse.org/

[6] Data Protection Commission (Ireland). (2020, February). *Guidance for Controllers on Data Security*. https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Data_Security_Guidance_Feb20.pdf

[7] *Irish Social Science Data Archive (ISSDA)*. (n.d.). University College Dublin. https://www.ucd.ie/issda/

[8] Slavkovic, A., Kinney, S., & Karr, A. (2013, August 2). O Privacy, Where Art Thou? *Chance*, *24*(4), 41-45. https://doi.org/10.1080/09332480.2011.10739886

[9] *National Opinion Research Center*. (n.d.). NORC at the University of Chicago. https://www.norc.org

[10] Thissen, M. R., & Mason, K. M. (2019, April 15). Planning security architecture for health survey data storage and access. *Health Systems*, *9*(1), 57-63. https://doi.org/10.1080/20476965.2019.1599702

Horton, Perry, and Bishop (2020)[11] present three tiers: (1) Open, (2) Accountable and (3) Controlled. The term restricted applies to both accountable and controlled. In our view, three levels are not sufficient for providing access to research data.

In "Sharing Confidential Data for Research Purposes A Primer", Reitera and Kinney (2011)[12] identify two primary restricted-access methods employed by most data stewards, including government agencies and individual investigators: licensing agreements and restricted-data centers. These access methods correspond to some of the tiers that we discuss below. Reitera and Kinney do acknowledge online enclaves such as NORC's for providing access, but do not specifically discuss other tiers.

Any discipline that analyzes research data is concerned with security. In addition to social sciences, medical, public health experts and epidemiologists are also dealing with how to provide appropriate access to research data. Lawyers and computing professionals tend to approach access to research data as a problem of licenses and waivers. DUAs and licenses in many circumstances are synonymous. A valid DUA or license is required to access the research data. While licenses concern who can access data and for how long, they do not specify a hierarchy of controls. Waivers are another way of saying unrestricted access. Computing professionals often focus on physical security, authentication, authorization, audit, and encryption. Training is often the specified human control. In our paradigm, the 10 security controls that we identify serve as extra precautions to ensure that disclosure information is not released.

Some paradigms treat "trustworthiness" as a continuum instead of as a minimum requirement. In our proposed approach, researchers must meet minimum requirements to access restricted data. A higher trust score does not entitle the researcher to relax security protocols nor automatically access other restricted-use data. Each restricted-use dataset requires a separate application.

In this paper, we propose seven tiers of access to research data. Each tier adds requirements that are necessary to mitigate disclosure risk and confirms appropriate management of the data. Improper handling of the data includes attempting to find a specific individual or household or failing to follow disclosure protection rules for data and output included in papers and presentations. By establishing a ladder of access conditions, each higher tier meets and exceeds the requirements of the lower tiers. While the highest tier meets all requirements, this tier will impede legitimate research for most data. The challenge for repositories is to provide access in a manner that promotes research while specifying security that provides appropriate protections against the risks of re-identification and harm. The tiers operationalize risk management options. The requirements of the research data determine the appropriate tier.

[11] *Open where possible, closed if necessary: reforming access categories for social science data archives* [Presentation]. (2020, February 17). International Digital Curation Conference 2020 (IDCC20), Dublin, Ireland. https://doi.org/10.5281/zenodo.3670943

[12] Reiter, J. P., & Kinney, S. K. (2011, September). Commentary: Sharing Confidential Data for Research Purposes: A Primer. *Epidemiology*, *22*(5), 632-635. https://doi.org/10.1097/EDE.0b013e318225c44b

Researchers must qualify for access in that level, and all access is through that tier or a more restricted tier only.

The tiers of access range from 0-Unrestricted to 6-Batch. At all tiers, research data are not to be used for identifying individuals or organizations. 0-Unrestricted (public access) does not have any special requirements while 6-Batch has all controls. The seven tiers are:

- 0-Unrestricted
- 1-Registered
- 2-Approved
- 3-Local
- 4-Remote
- 5-Coldroom
- 6-Batch

Tiers 3 through 6 are typically grouped together under the classification of restricted; these tiers require an application and Data Use Agreement (DUA). Tiers 1 and 2 fill the gap between unrestricted and restricted in situations where researchers still need to apply for access, but do not need a DUA.

The seven tiers build on 10 controls for accessing research data. These 10 controls protect against the disclosure, re-identification, and harm risks associated with a particular dataset. These regulations form a ladder and allow a tier to build on the protocols of lower tiers.

- Application
- Approval
- Agreement
- Period of Access
- Research Location
- Encryption
- Internet
- Output
- Proctor
- View Data

The table at the end of this paper shows how the access tiers mesh with the controls. At each level, researchers must agree and comply with all regulations. As risks increase, researcher agreement is not sufficient; technical configurations must prevent researchers from inadvertent data disclosure. While researchers agree to follow all conditions, each tier adds a layer of security that ensures researcher compliance. These extra security layers are an impediment to research and should be only implemented when risks of re-identification and harm necessitate.

More details about each security control follow:

- **Application** to analyze the data. Only data in the 0-Unrestricted tier do not require an application. For data in tiers 3-Local and above, researchers must submit IRB approval, a security plan, and confidentiality pledges. Furthermore, for data in tiers 3 and above, only researchers who can serve as Principal Investigators (PI) may apply. Those researchers who are not PI eligible, such as graduate students, may analyze the data only under the supervision of a PI.
- **Approval** to analyze the data. While tier 1-Registered-use requires researchers to submit information about research plans, only tiers 2 (Approved) and above require approval before access. If approval is required, researchers must wait after applying before gaining access.
- **Agreement** is whether the researcher only or researcher and an institutional (or university or organizational) representative must sign the DUA to access the restricted-use data. At tier 2-Approved and below, researchers can obtain access through their own agreement. For tier 3-Local and above, a university or institutional representative with authority to obligate the researcher’s organization as well as the researcher must agree to and sign the DUA.
- **Period of Access** is either unlimited, limited, or only for a specified period. Tier 3-Local and above are only accessible until an end date. 2-Approved may have limits on how long researchers can access the data. Agreements that require a university of institutional signature are always time bound.
- **Research Location** is where the data will be viewed. The research office has the client computer which may store the data or be a portal to a server where the data are stored. For tier 2-Approved and above, the client location must be private; accessing the data from a library or café is not permitted. A private location prevents inadvertent eavesdropping of the computer screen.
- **Encryption** at rest and in transit. Tiers 2-Approved and above require encryption. Encryption alleviates the ramifications of theft, loss of data, or interception. Research data that require approval (tiers 2 and above) must implement encryption in transit and at rest.
- **Internet** concerns both inbound and outbound network traffic. For tier 3-Local and above, access to the internet must be blocked. For tier 4-Remote, while the server allows inbound session connections only; outbound connections and other types of inbound connections are not allowed. The purpose of blocking the internet is to prevent researchers from inadvertently copying files to unauthorized locations (typically through drag and drop). An acceptable configuration implements a two-step process of copying files off the computer with the research data. Blocking the internet also prevents the computer from being compromised and having any file stolen.
- **Output** must be vetted for compliance with disclosure protection rules such as minimum cell counts and minimum subsample sizes for regressions. Tier 3-Local and above require vetting, but all levels require compliance with rules about output. While tier 3 authorizes self-vetting, tier 4-Remote and above require trained personnel who

are not part of the research project to review files before release in addition to the project team.

- **Proctor** monitors researchers while accessing the data. For all tiers, researchers are not allowed to look up specific respondents in the data nor transcribe data points. For Tier 5-Coldroom and above, researchers may only access the data in the presence of a proctor. Tier 5 and above prevent unauthorized use of the data through monitoring.
- **View Data** controls whether researchers can access the micro data as well as summary results. At Tier 6-Batch, researchers cannot view the micro data. While at all tiers, researchers agree to not attempt to re-identify or look up a particular respondent, at Tier 6, researchers are prevented from even accessing the micro data, so any re-identification or lookups are impossible.

Seven Tiers

Let us consider the seven proposed tiers of access in more detail. All research data regardless of tier are for the calculation of summary measures only and must not be used to locate an individual, organization, or community. Higher tiers build on lower tiers by adding more security controls.

0-Unrestricted

Public-access research data are typically available for download from websites. These data are available without restrictions on access. Disclosure and harm risks are negligible; nevertheless, the data are for research only and must not be used to locate an individual, organization, or community. Unrestricted research data are also labeled public-use and open-data. In many situations, public-use research data may be downloaded anonymously. The researcher who downloads the data can agree to the terms of use.

- *Example study:* Baby's First Years[13] (public) has data of this type.
- *Implementation:* Website and bandwidth to handle download demand.
- *Weakness:* Data might still have hidden risks.
- *Impediment to research:* Data may not contain sufficient information for analysis.

1-Registered

Registered research data are also typically available for download from websites. Disclosure and harm risks are very low. Unlike public-access data, registered data may not be downloaded anonymously. To register, researchers must provide valid contact information and a research purpose; however, download of the data does not require approval. The researcher who downloads the data can agree to the terms of use.

- *Example study:* National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018 [Public Use][14] requires registration from all data analyzers.

[13] Magnuson, Katherine A., Noble, Kimberly, Duncan, Greg J., Fox, Nathan A., Gennetian, Lisa A., Yoshikawa, Hirokazu, and Halpern-Meekin, Sarah. Baby's First Years (BFY), New York City, New Orleans, Omaha, and Twin Cities, 2018-2019. Inter-university Consortium for Political and Social Research [distributor], 2020-11-16. https://doi.org/10.3886/ICPSR37871.v2

[14] Harris, Kathleen Mullan, and Udry, J. Richard. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018 [Public Use]. Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor], 2022-02-09. https://doi.org/10.3886/ICPSR21600.v24

- *Implementation:* Registration system to collect information. Website and bandwidth to handle download demand.
- *Weakness:* Researchers could provide inaccurate information.
- *Impediment to research:* Researcher must provide information to access data.

2-Approved

Approved research data require registration and approval before download. While these data have very low disclosure risk, they may contain information that could be construed as having sensitivity. Because the data are only available upon approval, researchers may have to implement additional safeguards for these data such as encryption. The researcher may only be allowed to access the data in a private setting. The researcher who downloads the data can agree to the terms of use. In some cases, a department chair or graduate advisor may need to supervise the research.

- *Example studies:* Some Panel Study of Income Dynamics (PSID)[15] and Health and Retirement Study (HRS)[16] data fall into this category.
- *Implementation:* Application system with encrypted download.
- *Weakness:* Researchers could leak data inadvertently.
- *Impediment to research:* Researchers must apply for access to research data.

3-Local

Research data in this tier are restricted; however, access to the data is at the researcher's local university or organization. These data have a higher risk of re-identification and harm if disclosure occurs. Local data require an application and approval, but unlike 2-Approved, an institutional representative must sign the DUA in addition to the researcher. In the DUA, the institutional or organizational representative must verify that the researcher is qualified and affiliated with the institution. Moreover, the institution must have rules governing research misconduct and must agree to invoke these protocols if an infraction occurs. Qualified researchers must be PI eligible to access these data; other researchers and students must work under the supervision of a qualified researcher. Analysis of these data requires IRB approval and confidentiality pledges from personnel who can access the data. In addition to whole disk encryption, the data must reside on a standalone (non-networked) computer in a private office. The researcher must also agree to abide by disclosure protection rules and must self-review articles and output for compliance.

- *Example study:* NICHD Study of Early Child Care and Youth Development (SECCYD)[17] has data in this tier.
- *Implementation:* Standalone (non-networked computer) in a locked private office. Some organizations may have an acceptable server set up.

[15] Johnson, David S., Freedman, Vicki A., Sastry, Narayan, McGonagle, Katherine A., Brown, Charles, Fomby, Paula, … Stafford, Frank P. Panel Study of Income Dynamics (PSID): Main Interview, 1968-2015. Inter-university Consortium for Political and Social Research [distributor], 2018-10-04. https://doi.org/10.3886/ICPSR37142.v1

[16] *Health and Retirement Study, public use dataset*. (n.d.). Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). https://hrsdata.isr.umich.edu/data-products/public-survey-data

[17] United States Department of Health and Human Services. National Institutes of Health. Eunice Kennedy Shriver National Institute of Child Health and Human Development. (n.d.). *NICHD Study of Early Child Care and Youth Development: Phases I-IV [United States]*. Inter-university Consortium for Political and Social Research [distributor]. https://www.icpsr.umich.edu/web/DSDR/series/233

- *Weakness:* The research data with re-identification and harm risks are not under the control of the repository. Unauthorized access is possible.
- *Impediment to research:* Difficult to collaborate with a research team. Universities and organizations may be reluctant to permit a non-networked computer. Researchers may not have extra funds to buy a second computer that is dedicated to a single project.

4-Remote

Research data in tier 4-Remote have the same application requirements as the previous level, 3-Local. Instead of researchers analyzing the data on systems at the local organization, they access the data through encrypted connections to a "Virtual Data Enclave" or "Virtual Research Data Center". These data may have higher re-identification and harm risk. In some cases, these data may be linked with other information such as geographic contextual variables. These data are stored in an enclave and cannot be downloaded to the local computer, so that the repository retains control over access to the data. Researchers must review their output for compliance with disclosure protection rules; however, trained staff who are not members of research projects must also vet the output. Only files that meet disclosure protection requirements are released out of the enclave. Restrictions on additional data to be analyzed can also be enforced. Besides allowing a secondary level of output vetting, enclaves offer the additional benefit of enabling research teams to collaborate on the analysis of data with disclosure risk. Enclaves are fast becoming the preferred method for restricted data access and eventually will subsume Tier 3-Local.

- *Example study:* The restricted Los Angeles Family and Neighborhood Survey (L.A.FANS)[18] data are in this tier.
- *Implementation:* Terminal Server or Virtual Desktop Infrastructure (VDI) that prevents files from being copied off the server or VDI. ICPSR[19], NORC[20], and Survey Research Center[21] at the Institute for Social Research[22] have enclaves in production.
- *Weakness:* Researchers could still transcribe information from the screen.
- *Impediment to research:* Researchers must wait for the release of results. Available software may be limited. The computation power of the virtual machines may not be sufficient for some research.

5-Coldroom

Coldroom protocols add a proctor to the security requirements. The proctor checks that non-approved information is not extracted from the data. As with tiers 3-Local and 4-Remote, this level requires an application and approval. These data typically have even higher re-identification and harm risks.

[18] Pebley, A. R., & Sastry, N. (n.d.). *Los Angeles Family and Neighborhood Survey (L.A.FANS), Waves 1-2: Restricted Data Versions 1-3; Restricted Neighborhood Observations Data*. Inter-university Consortium for Political and Social Research [distributor]. https://www.icpsr.umich.edu/web/DSDR/series/846

[19] *Inter-university Consortium for Political and Social Research*. (n.d.). ICPSR. http://icpsr.umich.edu

[20] *National Opinion Research Center*. (n.d.). NORC at the University of Chicago. https://www.norc.org

[21] *Survey Research Center*. (n.d.). https://www.src.isr.umich.edu/

[22] *Institute for Social Research*. (n.d.). https://isr.umich.edu/

- *Example study:* Videos are an example of data in this tier. Most videos are high disclosure risk. As with data in 4-Remote, all output and files are reviewed before release.
- *Implementation:* Locked room with proctor. Census Research Data Centers[23] have implemented this level of security.
- *Weakness:* Researchers could still look up an individual record.
- *Impediment to research:* Accessing the data requires travel to the coldroom and an appointment. Repository staff must also allocate time to work in the coldroom to serve as proctors.

6-Batch

This tier is for research data with the highest risks and provides the maximum protections since researchers are unable to view the data. Researchers are only allowed to see approved summary results and are not able to view the micro data. Accessing these data requires an application and approval as well as institutional agreement. While this tier does not allow researchers to touch the micro data, this level has one advantage over 5-Coldroom in that it does not require travel.

- *Example:* Data with high sensitivity and high re-identification risks.
- *Implementation:* Batch system. A server with synthetic data and the software available in the batch system for testing programs will enable the system to run smoothly. *LISSY* at the Cross-national Data Center in Luxembourg[24] is an implementation of this tier. The retired *ANDRE* system at the National Center for Health Statistics (NCHS)[25] was also an example.
- *Impediment to research:* Without access to the data, analysis is cumbersome and requires much more time.
- Even though 6-Batch is more restrictive than 5-Coldroom, the tier does not require travel to a specific location.

In conclusion, while tiered access to research data is not a new idea, more than two or three levels are needed to meet the diverse needs of the research community. In this paper, we propose seven tiers along with detailed descriptions of each tier as well as examples of data that fall within each tier. With these seven tiers of access, repositories can meet the needs of researchers while still providing appropriate protections for research data. The tiered approach enables repositories to require sufficient security controls without creating unnecessary impediments to research.

Bibliography

[23] *Federal Statistical Research Data Centers*. (n.d.). Census Bureau. https://www.census.gov/about/adrm/fsrdc.html

[24] *LIS Cross-National Data Center in Luxembourg*. (n.d.). https://www.lisdatacenter.org/data-access/lissy/

[25] *NCHS - National Center for Health Statistics*. (n.d.). CDC. https://www.cdc.gov/nchs/index.htm

Crosas, M. (n.d.). *CIO Review: Cloud Dataverse: A Data Repository Platform for the Cloud*. https://openstack.cioreview.com/cxoinsight/cloud-dataverse-a-data-repository-platform-for-the-cloud-nid-24199-cid-120.html

Data Protection Commission (Ireland). (2020, February). *Guidance for Controllers on Data Security*. https://www.dataprotection.ie/sites/default/files/uploads/2020-04/Data_Security_Guidance_Feb20.pdf

Desai, T., Ritchie, F., & Welpton, R. (n.d.). Five Safes: designing data access for research. In *Economics Working Paper Series 1601*. University of the West of England, Bristol. https://www2.uwe.ac.uk/faculties/bbs/Documents/1601.pdf

Federal Committee on Statistical Methodology: Confidentiality and Data Access Committee. (2002, April). *Restricted Access Procedures*. https://nces.ed.gov/FCSM/pdf/CDAC_RAP.pdf

Federal Statistical Research Data Centers. (n.d.). Census Bureau. https://www.census.gov/about/adrm/fsrdc.html

Harris, Kathleen Mullan, and Udry, J. Richard. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018 [Public Use]. Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor], 2022-02-09. https://doi.org/10.3886/ICPSR21600.v24

Health and Retirement Study, public use dataset. (n.d.). Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). https://hrsdata.isr.umich.edu/data-products/public-survey-data

Institute for Social Research. (n.d.). https://isr.umich.edu/

Inter-university Consortium for Political and Social Research. (n.d.). ICPSR. http://icpsr.umich.edu

Irish Social Science Data Archive (ISSDA). (n.d.). University College Dublin. https://www.ucd.ie/issda/

Johnson, David S., Freedman, Vicki A., Sastry, Narayan, McGonagle, Katherine A., Brown, Charles, Fomby, Paula, … Stafford, Frank P. Panel Study of Income Dynamics (PSID): Main Interview, 1968-2015. Inter-university Consortium for Political and Social Research [distributor], 2018-10-04. https://doi.org/10.3886/ICPSR37142.v1

Kinney, S. K., Karr, A. F., & Gonzalez Jr., J. F. (2010). Data Confidentiality: The Next Five Years Summary and Guide to Papers. *Journal of Privacy and Confidentiality*, *1*(2). https://doi.org/10.29012/jpc.v1i2.569

LIS Cross-National Data Center in Luxembourg. (n.d.). https://www.lisdatacenter.org/data-access/lissy/

Magnuson, Katherine A., Noble, Kimberly, Duncan, Greg J., Fox, Nathan A., Gennetian, Lisa A., Yoshikawa, Hirokazu, and Halpern-Meekin, Sarah. Baby’s First Years (BFY), New York City, New Orleans, Omaha, and Twin Cities, 2018-2019. Inter-university Consortium for Political and Social Research [distributor], 2020-11-16. https://doi.org/10.3886/ICPSR37871.v2

National Opinion Research Center. (n.d.). NORC at the University of Chicago. https://www.norc.org

NCHS - National Center for Health Statistics. (n.d.). CDC. https://www.cdc.gov/nchs/index.htm

Open where possible, closed if necessary: reforming access categories for social science data archives [Presentation]. (2020, February 17). International Digital Curation Conference 2020 (IDCC20), Dublin, Ireland. https://doi.org/10.5281/zenodo.3670943

Pebley, A. R., & Sastry, N. (n.d.). *Los Angeles Family and Neighborhood Survey (L.A.FANS), Waves 1-2: Restricted Data Versions 1-3; Restricted Neighborhood Observations Data*. Inter-university Consortium for Political and Social Research [distributor]. https://www.icpsr.umich.edu/web/DSDR/series/846

Reiter, J. P., & Kinney, S. K. (2011, September). Commentary: Sharing Confidential Data for Research Purposes: A Primer. *Epidemiology*, *22*(5), 632-635. https://doi.org/10.1097/EDE.0b013e318225c44b

Slavkovic, A., Kinney, S., & Karr, A. (2013, August 2). O Privacy, Where Art Thou? *Chance*, *24*(4), 41-45. https://doi.org/10.1080/09332480.2011.10739886

Survey Research Center. (n.d.). https://www.src.isr.umich.edu/

The Dataverse Project. (n.d.). https://dataverse.org/

Thissen, M. R., & Mason, K. M. (2019, April 15). Planning security architecture for health survey data storage and access. *Health Systems*, *9*(1), 57-63. https://doi.org/10.1080/20476965.2019.1599702

United States Department of Health and Human Services. National Institutes of Health. Eunice Kennedy Shriver National Institute of Child Health and Human Development. (n.d.). *NICHD Study of Early Child Care and Youth Development: Phases I-IV [United States]*. Inter-university Consortium for Political and Social Research [distributor]. https://www.icpsr.umich.edu/web/DSDR/series/233

Access Tier by Control

	Tier	Description	Application	Approval	Agreement	Period of Access	Research Location	Encryption	Internet	Output	Proctor	View Data
Public	0-Unrestricted	researcher may download	none	none	Researcher	No limit	public or private	not required	allowed	not vetted	not monitored	allowed
	1-Registered	researcher must provide additional info such as research purpose before download	submit information	none	Researcher	No limit	public or private	not required	allowed	not vetted	not monitored	allowed
	2-Approved	researcher must be approved before download	must apply	approved	Researcher & Advisor	Limited	private	at rest in transit	allowed	not vetted	not monitored	allowed
Restricted	3- Local	researcher receives data with approved security plan	must apply	approved	Researcher & Institution	Specified period	private	at rest, real-time in transit	blocked	self-vetted	not monitored	allowed
	4-Remote	researcher comes to data electronically with approved security plan	must apply	approved	Researcher & Institution	Specified period	private	at rest in transit	blocked except session	externally vetted	not monitored	allowed
	5-Coldroom	researcher comes to data in person with pre-approved materials	must apply	approved	Researcher & Institution	Specified Period	private	at rest in transit	blocked	externally vetted	watched during access	allowed
	6-Batch	researchers cannot access the data researchers can only access summary results	must apply	approved	Researcher & Institution	Specified period	private	at rest	only batch submissions	externally vetted	monitored batch jobs	not allowed