CHAPTER 3

Model Data Use Agreements: A Practical Guide

Amy O'Hara (Georgetown University)

3.1 Overview

What are data use agreements? Data use agreements (DUA)—also referred to as data sharing agreements or data use licenses—are documents that describe what data are being shared, for what purpose, for how long, and any access restrictions or security protocols that must be followed by the recipient of the data. Other contracts, such as non-disclosure agreements, may be used to guarantee confidentiality over sensitive discussions, information, and data.

This chapter explains how to develop a DUA to access administrative data for a research project. The chapter documents specific questions to consider when developing an agreement and points to useful templates and guides.

There are at least two parties to such agreements: the data provider and the data requestor. The data provider is responsible for permitting data access on behalf of the collecting agency or data subjects.

Cite as: O'Hara, Amy. "Model Data Use Agreements: A Practical Guide." In: Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. 2020.

> What if the data provider does not require any formal documentation? The researcher should write a letter describing the data requested, the planned uses, and a summary of the data management plan. The letter should clearly state the proposed use of the data, redistribution of the data, and methods for data retention or destruction at the project's end. Researcher and data provider should then sign and date the letter. Alternatively, the researcher can simply send the letter and obtain a return receipt.

The data provider is bound by law, regulation, or policies that may be very specific regarding access to direct identifiers (name, date of birth, social security number) and sensitive information (health conditions, grades, or test scores). The data requestor is a researcher pursuing data access for a specific purpose. Researchers at universities must typically go through a review of the DUA by an Office of Research or Sponsored Programs or the Office of the General Counsel and possibly by university information security specialists.

In some circumstances, the data provider may utilize a separate data custodian or data intermediary to offer data on their behalf, adhering to all required laws, regulations, and policies. Custodians and intermediaries support data access, reducing the burden for data providers by handling requests, reviews, and provisioning to researchers. Projects involving multiple information sources will require multiple DUAs, potentially involving a variety of terms and conditions. DUAs may also become more complex for multi-site research projects when different teams of researchers will need to access data and collaborate. Intermediaries can be particularly useful in these circumstances for facilitating data access, by coordinating between different data providers and researchers.

Depending on the data provider, other forms of documentation can be used. Examples include memoranda of understanding (MOU), data use agreements, and data exchange letters. These have different structures and levels of detail, but all of these instruments will state the legal framework for data access, what the requestor may do with the data (e.g., scope of the study, restrictions on redistribution), security

controls, and constraints on publishing. The data requestor should always prepare some form of documentation for data access, even if the data provider does not require it.

3.1.1 Relating the DUA to the Five Safes Framework

The Five Safes framework used throughout this handbook is an approach for structuring aspects of data access. The five safes are safe projects, safe people, safe settings, safe data, and safe outputs.[1]

Safe projects have governance measures over project scope and sensitivity with review and approval processes that involve institutional review boards (IRB) or ethics boards. Data providers must determine who are *safe people* through policies, screening, and training, and may require affiliation to an educational or non-profit institution, proof of research competence (e.g., grants received, curriculum vitae), and citizenship or tenure in the relevant country. Safe settings and data involve the researcher's interface and work environment, potentially restricting what an analyst can see, what an analyst can do, the analyst's computing environment, and the analyst's physical location (see also chapter 2). Safe data and outputs protect the privacy of data subjects by reducing re-identification risks both during access and after publication. Such protection occurs through statistical disclosure limitation methods such as rounding, aggregating, and suppression (obscuring unique observations in tables, figures, or maps) or formal, mathematical privacy protections (see chapters 5 and 6).

At a high level, a DUA should address all five safes. It should include intended data uses to define the safe project; terms for data access and handling for a safe setting; and terms for output publication and release for safe outputs. DUAs are essential to define acceptable data uses, linkages, and topics of analysis. Agreements may also detail roles and responsibilities for the data provider and researchers (defining safe people) and cover safe data by including a list of data elements and any reporting or disposition requirements. There are many permutations

[1] See Desai, Ritchie and Welpton (2016) for more information on the Five Safes framework including examples for each dimension.

on such restrictions;[2] any requirements as well as penalties for failing to comply with them should be included in the DUA.

Such an agreement strives to protect all parties by specifying the terms and conditions for data access and use. DUAs are risk mitigation tools, clarifying expectations between the parties. Data providers are often reluctant to enter data sharing arrangements, as they may be fearful of the liabilities resulting from use of the data that could result in harm to their program, agency, or the data subjects. Through DUAs, data providers can specify controls on data handling and notification measures in case of data mismanagement. DUAs also solidify the roles and responsibilities of researchers and their institutions, clarifying liability issues in advance.

The following sections describe how to (1) prepare for a data sharing arrangement, (2) negotiate a sound agreement, and (3) comply with the signed agreement, based on review of guides and best practices across multiple domains.[3] Some of these refer to a researcher negotiating a DUA with a data provider for the first time, but the considerations for this case contain pointers for establishing good processes and developing templates and examples for subsequent DUAs.

3.1.2 Preparation

Creating DUAs can be time-intensive. In some cases, negotiations fall apart after months or years of discussions. Advance planning can help both researchers and data providers achieve sound DUAs. DUAs can be initiated by the researcher or data provider.[4] Data providers may have different or expedited procedures when sharing data with a researcher, an evaluator, or contractor working on their behalf.

If a data provider has an established data request process, a researcher must review their terms and requirements, offering additions or edits

[2] See Goroff, Polonetsky and Tene (2018) for a comprehensive discussion of possible methods.

[3] See Appendix B for a set of these guides.

[4] See Yates et al. (2018) for a checklist from the data provider's perspective.

as appropriate. Data providers should be aware of the laws, regulations, and policies permitting use of their data, and, upon receiving a first request, determine whether data request procedures already exist in their organization. Data providers (such as government agencies or private companies) may have Offices of General Counsel that have preferred templates or formats. Some data providers will be reluctant or unable to modify their request processes. Data request and access procedures may not always be publicly available, though some agencies and organizations have data request procedures on their websites, and this can significantly speed up and simplify the request process.

3.1.3 Understanding the Available Data

Researchers need to be able to identify the correct data source: the agency or organization who holds the data content needed for their planned analysis. This may be difficult in settings where data descriptions are not readily available. Can data users determine whether the data are fit for use? Can they ascertain what data is captured by data providers, how the data are coded, and whether such capture and coding are documented consistently across time?

Well-prepared data users will typically do this by reviewing a data description, a codebook, or a data dictionary. Data providers should consider preparing such materials or working with pilot data users to do so. A data sample may provide a better understanding of the data content. If documentation or a sample is unavailable, program rules, regulations, and forms can be used to provide background.

However, a field on an application or benefits form does not automatically mean the information is cleaned or stored by the agency. Prior analyses of the same data by other studies or at other sites can provide helpful information on availability and usability of the underlying data. Researchers should seek out such studies and providers may want to keep a record of research conducted with their data to facilitate future use.

3.1.4 Understanding the Costs of Obtaining Data

Both parties should consider what is possible, and what is likely, in terms of the timeframe the agreement will cover. This includes when data delivery can occur, how data will be extracted from administrative systems, and what expenses might arise during the term of the data sharing arrangement. Agreements can take up to a year to negotiate from drafting to execution, especially if there is no history of the two parties exchanging data before. Even organizations with past data sharing relationships or with established processes may have a queue of requests, which may create delays. After achieving a signed agreement, researchers should anticipate for the time between approval and delivery: the processes for fulfilling the request may be intensive. For example, data providers will need time to document and format the requested data and additional time may be needed to pull data from multiple databases or from inactive storage. That process may be especially lengthy if the request is novel. Data providers may also require notification or approvals before any output releases or publications.

Many administrative agencies are resource constrained, needing to prioritize program needs over research requests. In this situation, they may decide to charge fees for data preparation and extraction. Being transparent about timeframe and cost and making the data use agreement as clear as possible helps set expectations between the parties.

3.1.5 Consideration for the Data Subjects

Researchers should consider potential benefits, costs, and risks for the data subjects in the planned project and think of how to communicate the project to the data subjects, including an explanation of why their data are needed. The researchers should be prepared to explain what data will be used, whether the data will be linked with other information, and who will have access to the data. They should also be able to explain the project in direct language (free from jargon) for the subjects or their parents or guardians and provide a finite project timeline. This is useful for purposes of establishing an informed consent

Researchers may consider preparing (and data providers may consider requesting) an engagement matrix that maps project steps with different forms of external input to build trust with the data subjects (Future of Privacy Forum and Actionable Intelligence for Social Policy, 2018). Engagement could involve simply informing subjects about the project, seeking their input, or active collaboration during the project. Communicating with the subjects could include interviews, advisory committees, working groups, town halls, social media discussions, or press releases. Researcher and data provider may also consider a transparency checklist[a] as part of each project,[b] to add legitimacy to the project and its results when completed. A transparency checklist can accompany publications resulting from the analysis to clarify how the data, code, and other study materials were handled upon project completion.

[a] http://www.stat.columbia.edu/~gelman/research/published/checklist.pdf (accessed 2020-12-15).

[b] See Aczel et al. (2020) guide and checklist.

procedure as well as the conduct of ethical research when consent is not required and for communication with the public (e.g., in contexts where the research informs public policy). The ethical and transparent conduct of research supports future use of the data and establishes trust with the public and data subjects.

3.1.6 Investigating the Data Sharing History for Data Providers and Researchers

Researchers might inquire whether the data needed for the project have been successfully shared by the data provider before. In relevant cases it can be helpful to build on a copy of the previous data use agreement, provided by the agency or by researchers who have accessed data in the past.[5] For a researcher, requesting data access with a past protocol in hand is a strong position. When approaching

[5] Some jurisdictions may require a formal written request or even a Freedom of Information Act request to share the DUAs.

an agency with a set process for data sharing, the researcher should review the process and forms and know which office in the organization approves requests. If requesting an unusual extract or approaching an agency that has never permitted research access before, researchers should identify some data sharing examples within their department or in other localities to review terms and conditions in their agreements. Data providers on the other hand can ask researchers about past performance information on quantitative research projects. This could include their history of using administrative data or examples of their data management plans and approaches when handling sensitive data. This information can help the data provider determine whether the researcher has the capacity to protect the data, deliver the results they have proposed, and whether they have been good partners in the past (or whether they have been involved with data breaches).

3.1.7 Understanding the Legal Context

It is important to have an understanding of the legal framework that governs the use of the data. This may involve laws at the national, subnational (state, province), and local level. In the case of private data providers, it may involve notions of copyright and legal responsibility. If the data provider and the research institution are not located in the same country, this includes the legal framework in both countries. If the server hosting the data is based in a third country, additional requirements may affect the data provider (e.g., the General Data Protection Regulation (GDPR) in the European Union). The degree of regulation varies across countries, and data protection laws (and interpretations of them) change frequently. The parties should work with legal and privacy professionals to identify the legal authority for data access. This is especially important when requesting individually identified data, as defining what constitutes personal data varies across jurisdictions.

Investigating the legal framework helps researchers form realistic expectations regarding scope and conditions for the DUA. Moreover, it is important that researchers (or their institutions) are aware of the legal

setting, so they can ensure compliance with all applicable laws, especially if the data provider has limited legal experience approving data sharing and data use by researchers.

3.1.8 Thinking through the Analysis and Publication Process

Considering the project goals and timeline, the researcher should assess how much time it will take to clean, harmonize, and link data—all necessary steps before conducting analyses or publishing results. Time required for each of these steps can depend on the past experiences of the researcher (or their institution) with a particular type of data. Researchers should allow ample time to prepare data for use after receipt, possibly in collaboration with the data provider. The researcher should also allocate time to prepare findings for release and identify disclosure avoidance techniques to protect against re-identification of the data subjects in project outputs. Data providers should be prepared to review outputs and be familiar with common disclosure avoidance protocols (see chapter 5).

3.1.9 Taking a Broad Interpretation of Data

Data includes information directly from administrative databases on program participants or clients, regardless of the extent to which it is processed, linked, or contains identifiers. But data also refers to metadata about the system, files, and content as well as statistical information that will be published through the project, such as descriptive statistics, coefficients, or visualizations. A sound data use agreement covers all of these. See the concepts of safe data and safe outputs in section 3.1.1 on relating the DUA to the five safes framework.

3.2 Negotiating the Data Use Request

With preparations complete, the data provider and researchers can pursue a DUA for an individual project. The data provider ultimately decides whether and how access will be granted: a researcher with clear plans and expectations and a data provider with established and transparent processes are equipped to engage effectively. This section includes some pointers and considerations for the pursuit of a DUA by a researcher, especially in a first-time engagement. From the provider perspective, many of the points below are about information the researcher needs, and data providers can facilitate the DUA process by making this information available either publicly or to the individual researcher. Data providers may also face similar issues if they are requesting data from other agencies or organizations.

3.2.1 Getting the Right People Involved

The researcher needs to communicate with the right decision-makers within the data providing organization about the project and upcoming request. Note that administrators may support the idea of the project but may be unaware that their data systems lack necessary data elements to complete the analysis. An administrator might not have a full view of the complexities of their data systems and structures, which may make it difficult or impossible to identify or derive the data needed for the analysis without technical assistance. Similarly, substantial resources from the data provider may be required to extract data from multiple systems and, if a longitudinal study is planned, from active and inactive storage. It is therefore important to consult the data provider's technical staff on each request. Researchers will need to engage their Office of Sponsored Research, IRB, and sometimes Office of General Counsel. When working in a foreign country, many parties may need translations (even if the researcher does not).

3.2.2 Asking Questions About the Process

The researcher should discuss with the data provider how the negotiation will proceed before submitting the request. Does the data provider have an iterative process? Will they counter or iterate on the request? If one part of the request is denied, will the rest proceed or will the whole request be returned? Does the data provider require an IRB or ethics board review and approval from their end, or do they require that a researcher obtain IRB approval from their institution before requesting or accessing data? What is the signature process for all parties to the agreement? Who are authorized individuals permitted to sign on behalf of the researcher's or data provider's organization? Will the data provider require background checks on researchers?

3.2.3 Understanding the Reasons Behind a Negative Response

Data providers say no for many reasons. It is important to understand what the "no" means in order to determine how best to respond. The researcher should determine whether the response is stemming from a legal, policy, or cultural barrier.

Organizations without existing systems for data sharing may turn down a request because they lack clear internal roles and responsibilities or resources to administer the agreement development, data exchange, and relationship monitoring. Obtaining funding or external resources can help to support the process.

A request denial may also come from a key decision-maker who may feel that the risks of data sharing overwhelm potential benefits. They may have concerns about unauthorized uses, breaches, negative publicity, or privacy concerns raised by their legislatures or clients. Decision-makers may be afraid that problems will be discovered in the data or have trepidation about what the results of the study will show. Such concerns are described in "Why Data Providers Say No... and Why they Should Say Yes" (National Neighborhood Indicators Partnership, 2018). The engagement matrix and transparency check list,

described in the breakout box on communication tools for engaging subjects and the public, can help in this area.

If data are inaccessible due to a legal barrier, the researcher should find the section of the statute or code that prohibits access and determine whether access would be permitted in the case that the researcher were under contract with the agency or producing an output for that agency. In instances where access would have been permitted, the parties may consider discussing a mutually beneficial contractor relationship between the researcher and data provider. Otherwise, the researcher may determine whether a separate legal interpretation of the statute or regulation would be appropriate or whether the law effectively prohibits access. Even when there are not legal barriers, there may be policy barriers. This happens when a written policy prevents access. The parties should investigate whether a waiver or a policy change are feasible.

When there is no law or written policy blocking access, there still may be cultural barriers. Data providers (or individuals at the data provider) may reject a request because such sharing has never taken place before or was done only in special circumstances. They may also lack the resources to entertain the request: they may have already shared the data with another research team or their own in-house experts are looking into the same or related research topic. The researcher can try to identify why the agency is reluctant and explore the risks that data sharing poses to them. They can discuss with the data provider how controls over the mode of access, users, uses, and outputs may mitigate these risks and how the project can produce benefits for the provider. Negotiating parties can refer to the various sections in this handbook for examples on successful data use agreements, as well as the technical possibilities (see chapters 2 and 5), which might allay fears and uncertainties.

3.2.4 Trying to Find Mutual Interests

It is helpful to think through the interests of the organization as well as the interests of individual decision-makers, such as the program

manager, agency leader, chief information security officer, and so on.[6] Consider what the agency needs to do: improve program administration, increase efficiency, reduce costs, and help program participants. What can the research team produce for the data provider? This could be clean data, documentation, code, a report, or a dashboard. Researchers should ask what the data provider's unanswered questions and needs are.

3.2.5 Drafting the Request

Does the agency have a posted process, pre-specified forms, or a template? If none exists, the researcher should try to get an example of a successful request and be attentive to detail in formulating a new request. Be sure to include processes and requirements of the data provider, such as review requirements.

Guides that provide templates are available from various domains. Appendix A to this chapter provides one template. Other examples are listed below:

- "Data Sharing: Creating Agreements" (Jarquín, 2012) from the Colorado Clinical and Translational Sciences Institute includes specific questions to help determine which sections should be included in a DUA from a clinical health perspective.
- *Legal Issues for IDS Use: Finding a Way Forward* (Petrila et al., 2017) is an expert panel report informing state and local governments that want to integrate data. This report explains why politics and relationships matter and walks through the legal considerations for preparing a MOU or Data Use License. The document includes links to a sample agreement made with two states and one county as well as a data license template from a federal agency for health and human services data.
- "Guidelines for Developing Data Sharing Agreements to Use State Administrative Data for Early Care and Education Research" (Shaw, Lin and Maxwell, 2018) includes examples with early childhood

[6] See Coburn, Penuel and Geil (2013) for a discussion of maintaining mutualism in a research partnership.

research from two states, along with links to checklists and toolkits. This research brief also includes “advice from researchers” sections throughout.

3.2.6 Signing the Agreement

Complications can arise during the signature process for agreements. Late edit insertions may require further rounds of review. When the document is signed by all parties (i.e., fully executed), both sides must monitor staffing changes in their organizations to keep the signatories current. Most agreements describe how changes to the executed agreement may be requested (e.g., in writing to the signatory, within fifteen days of a new appointment). If the researcher changes institutions, they must discuss the DUA update process with the original institution, new institution, and data provider so expectations are clear. Both the original signatory and the researcher should determine whether the original DUA will be terminated once a new DUA with the gaining institution is signed. The researcher must follow data management and security protocols if data transfer to their gaining institution is required, checking with institutional information security specialists if terms of transfer were not explicit in the original DUA.

3.3 Compliance

Once the agreement is signed, the work is not done. The researcher should develop a plan to ensure compliance with the terms in the agreement and implement measures to demonstrate compliance per DUA requirements. Monitoring data processing controls, lists of approved users, updates to storage locations, upcoming releases, and review of publications requires coordination across the research team. Even if the data provider is not tracking these things, the researcher should.

The researcher should review the agreement terms regularly to be sure the necessary data are accessible and the project is on track for completion within the stated scope and timeline. If the researcher discovers

a need for additional data elements, an extension, or broader scope, they need to pursue a modification to the agreement. Since such modifications are common, the data provider may consider developing a template.

When using the data, the researcher should remember that this is a contractual arrangement and an opportunity to build trust between the parties. Working collaboratively with the data provider to understand the data will help build this relationship. Administrative data were not originally collected for research use, so researchers should ask questions if the data do not look as expected. Seeking clarification or correction can avoid misuse of the data and keep the data provider involved.

3.4 Summary

No matter the size of the project or the volume of data needed, all parties should invest the time in preparing a sound data use agreement. Agreements enable safe projects. The topics covered in this chapter have been put in to practice through all the case studies in this volume. The process is well described in chapter 12 on the Stanford-San Francisco Unified School District Partnership. Appendix A provides a sample text for consideration when writing DUAs, and Appendix B lists additional toolkits and guides on the DUA process.

About the Author

Amy O'Hara is a Research Professor in the Massive Data Institute at the McCourt School of Public Policy at Georgetown University, and the Director of Georgetown's Federal Statistical Research Data Center. She also leads the Administrative Data Research Initiative at Georgetown, improving secure, responsible data access for research and evaluation. She was previously a senior executive at the US Census Bureau where she negotiated DUAs for federal, state, and local administrative data. Her current research focuses on population measurement, data governance, and record linkage. She received her PhD in Economics from the University of Notre Dame.

References in Chapter 3

Aczel, Balazs, Barnabas Szaszi, Alexandra Sarafoglou, Zoltan Kekecs, Šimon Kucharský, Daniel Benjamin, Christopher D. Chambers, Agneta Fisher, Andrew Gelman, Morton A. Gernsbacher, John P. Ioannidis, Eric Johnson, Kai Jonas, Stavroula Kousta, Scott O. Lilienfeld, D. Stephen Lindsay, Candice C. Morey, Marcus Munafò, Benjamin R. Newell, Harold Pashler, David R. Shanks, Daniel J. Simons, Jelte M. Wicherts, Dolores Albarracin, Nicole D. Anderson, John Antonakis, Hal R. Arkes, Mitja D. Back, George C. Banks, Christopher Beevers, Andrew A. Bennett, Wiebke Bleidorn, Ty W. Boyer, Cristina Cacciari, Alice S. Carter, Joseph Cesario, Charles Clifton, Ronán M. Conroy, Mike Cortese, Fiammetta Cosci, Nelson Cowan, Jarret Crawford, Eveline A. Crone, John Curtin, Randall Engle, Simon Farrell, Pasco Fearon, Mark Fichman, Willem Frankenhuis, Alexandra M. Freund, M. Gareth Gaskell, Roger Giner-Sorolla, Don P. Green, Robert L. Greene, Lisa L. Harlow, Fernando Hoces de la Guardia, Derek Isaacowitz, Janet Kolodner, Debra Lieberman, Gordon D. Logan, Wendy B. Mendes, Lea Moersdorf, Brendan Nyhan, Jeffrey Pollack, Christopher Sullivan, Simine Vazire, and Eric-Jan Wagenmakers. 2020. "A consensus-based transparency checklist." *Nature Human Behaviour*, 4(1): 4–6. https://doi.org/10.1038/s41562-019-0772-6.

Coburn, Cynthia E., William R. Penuel, and Kimberly E. Geil. 2013. "Research-Practice Partnerships: A Strategy for Leveraging Research for Educational Improvement in School Districts." William T. Grant Foundation. https://wtgrantfoundation.org/library/uploads/2015/10/Research-Practice-Partnerships-at-the-District-Level.pdf (accessed 2020-10-05).

Desai, Tanvi, Felix Ritchie, and Richard Welpton. 2016. "Five Safes: Designing data access for research." https://uwe-repository.worktribe.com/output/914745 (accessed 2020-01-30).

Future of Privacy Forum, and Actionable Intelligence for Social Policy. 2018. "Nothing to Hide: Tools for Talking (and Listening) About Data Privacy for Integrated Data Systems." https://fpf.org/wp-content/uploads/2018/09/FPF-AISP_Nothing-to-Hide.pdf.

Goroff, Daniel, Jules Polonetsky, and Omer Tene. 2018. "Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data." *The ANNALS of the American Academy of Political and Social Science*, 675(1): 46–66. https://doi.org/10.1177/0002716217742605.

Jarquín, Paige Backlund. 2012. "Data Sharing: Creating Agreements." Colorado Clinical and Translational Sciences Institute & Rocky Mountain Prevention Research Center. http://trailhead.institute/wp-content/uploads/2017/04/tips_for_creating_data_sharing_agreements_for_partnerships.pdf.

National Neighborhood Indicators Partnership. 2018. "Why Data Providers Say No...And Why They Should Say Yes." https://www.neighborhoodindicators.org/library/guides/why-data-providers-say-noand-why-they-should-say-yes (accessed 2020-07-15).

Petrila, John, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. 2017. "Legal Issues for IDS Use: Finding a Way Forward." University of Pennsylvania, Actionable Intelligence for Social Policy. https://www.aisp.upenn.edu/resource-article/legal-issues-for-ids-use-finding-a-way-forward/ (accessed 2020-10-05).

Shaw, Sara, Van-Kim Lin, and Kelly Maxwell. 2018. "Guidelines for Developing Data Sharing Agreements to Use State Administrative Data for Early Care and Education Research." Administration for Children & Families OPRE Research Brief 2018-67. https://eric.ed.gov/?id=ED602071 (accessed 2020-10-05).

Yates, Deborah, Tim Beale, Stewart Marshall, and Martin Parr. 2018. "Designing data sharing agreements: a checklist." Open Data Institute. https://gatesopenresearch.org/documents/2-44 (accessed 2020-10-05).

Appendix

Appendix A

Sample Text for Agreement Components

Often, simply establishing that a proposed agreement covers all the important components can be a major impediment. To assist with this, below is a list of agreement sections with example language sourced from a range of successful data use agreements; this is offered as a starting point, not legal advice.

Title
Data Use Agreement for [Data/System] Access between Party 1 and Party 2

Parties and Purpose
This Agreement is between Party 1 [Office, Agency, Department, Institution] and Party 2 [Office, Agency, Department, Institution]. Party 1 and Party 2 are entering into an Agreement that will allow the exchange of data and clarification of data access and use. Party 1 will provide data collected to Party 2 for the purposes of [specify].

Authority
Party 1 is a(n) [specify] organization whose mission is [specify]. The authority for Party 1 to enter into this Agreement is [xxx]. This authority permits the release of [data] to [specify]. The [law/code] permits disclosure of [data] for [specify] functions. Party 2 is an [specify] organization whose mission is [specify].

Terms and Conditions
Description of planned data use by Party 2, consistent with Purpose above.

- Treatment of data anomalies, including technical assistance from Party 1 and redelivery as needed
- Terms for data storage, treatment of original data, handling of Personally Identifiable Information, and data linkage protocols

- Conditions for storing modified data (including integrated, recoded, de-identified, and derived data) during and after the project
- Terms for storage of researcher generated files (including retention/archiving, e.g., To the extent permitted by law, the original data received from Party 1 will be retained by Party 2 for [specify period].)

Data Elements
The following data will be provided under this Agreement: [Specify list of data elements from named programs/systems, noting which time periods, populations, and/or geographies are sought.]

Approved Research Uses
[Describe project objectives, intended data use, expected linkages.]

Roles & Responsibilities
Party 1 agrees
To transfer to Party 2 via [specify, e.g., secure File Transfer Protocol or appropriately encrypted disk], data from [specify] for the years [specify], as described in [Data Elements]. The delivery of [specify] data will occur before [specify]. To disclose data only for the authorized uses in [Terms and Conditions]. To comply with all applicable federal and state laws and regulations relating to the use and disclosure, the safeguarding, confidentiality, and maintenance of the data. To provide adequate documentation and support of transferred files for Party 2 to be able to interpret the data for the uses permitted in this Agreement, including definitions of variables/data dictionary, a record layout, record count, and record length. To allow Party 2 to link with [specify] data to complete their analysis. To allow Party 2 to use the data at the Processing Sites listed in this Agreement for the projects listed in [Approved Research Uses] in this Agreement.

Party 2 agrees
To access, hold, use, and disclose data only for the authorized uses in [Terms and Conditions]. To comply with all applicable federal and state laws and regulations relating to the use and disclosure, the safeguarding, confidentiality, and maintenance of the data. To ensure that

all data users comply with the requirements of this Agreement. To immediately report within [specify] any use or disclosure of Protected Data other than as expressly allowed by this Agreement. Notice shall be given to the contact [specify]. Any changes in planned use of the data must be submitted to Party 1 in writing and receive written approval.

Duration, Amendments, and Modifications

This Agreement is effective on the date it is signed by both parties. The Agreement shall terminate [specify number of months/years] following the date on which it becomes effective. If, at the end of [same number of months/years above], the parties wish to continue the relationship, they must execute a new Agreement.

The parties shall review this Agreement at least once every [specify] or whenever a [State/Federal/Local] statute is enacted that materially affects the substance of the Agreement, in order to determine whether it should be revised, renewed or canceled.

Notwithstanding all other provisions of this Agreement, the Parties agree that

a. This Agreement may be amended at any time by written mutual consent of both parties and
b. Either party may terminate this Agreement upon thirty (30) days written notice to the other party.

Termination

Either party may terminate this Agreement for any reason on [specify number of days] business days' notice to the other party. Each party may terminate this Agreement with immediate effect by delivering notice of the termination to the other party, if the other party fails to perform, has made or makes any inaccuracy in, or otherwise materially breaches, any of its obligations, covenants, or representations, and the failure, inaccuracy, or breach continues for a period of [specify number of days] business days' after the injured party delivers notice to the breaching party reasonably detailing the breach.

Ownership of Developed Intellectual Property

If either party develops any new Intellectual Property in connection

with this Agreement, the parties shall enter into a separate definitive Agreement regarding the ownership of that new Intellectual Property.

Resolution of Disagreements
Should disagreement arise on the interpretation of the provisions of this Agreement, or its amendments and/or revisions, that cannot be resolved at the operating level, the area(s) of disagreement shall be stated in writing by each party and presented to the other party for consideration. If agreement on interpretation is not reached within thirty (30) days, the parties shall forward the written presentation of the disagreement to respective higher officials for appropriate resolution.

Confidentiality and Non-Disclosure
Party 2 shall use appropriate safeguards to protect the data from misuse and unauthorized access or disclosure, including maintaining adequate physical controls and password protections for any server or system on which the data is stored, ensuring that data is not stored on any mobile device (for example, a laptop or smartphone) or transmitted electronically unless encrypted, and taking any other measures reasonably necessary to prevent any use or disclosure of the data other than as allowed under this Agreement. Party 2 shall ensure that any agents, including subcontractors, to whom it provides the data agree to the same restrictions and conditions listed in this Agreement. Party 2 will not attempt to identify any person whose information is contained in any data or attempt to contact those persons.

IT Security
[Specify Statutes or Acts] protect the confidentiality of the data. Party 2 will comply with all laws applicable to the privacy or security of data received pursuant to this Agreement.

Publication/Disclosure Rules
Party 2 will ensure that any study, report, publication, or other disclosure of data provided under this Agreement is limited to the reporting of aggregate data and will not contain any information identifiable to a private person or entity. Aggregate data for purposes of this Agreement will mean datasets consisting of no fewer than [specify cell restrictions

or alternative disclosure limitation methods]. [Include citation and/or disclaimer language if desired.]

The dissemination and use of publicly released reports, articles, and other products derived in whole or in part from the data will not be discontinued due to the expiration or termination of this Agreement. Furthermore, the use of data linked to other data as part of the projects described in Attachment B will not be discontinued due to expiration or termination of this Agreement.

Party 2 agrees to provide Party 1 with an advance copy of any publication resulting from the data use not less than [specify number of days] prior to the submission or disclosure of the publication, to permit Party 1 to reasonably comment, update, or otherwise propose modifications or edits to the draft publication and to ensure there is no disclosure of confidential data. If Party 1 does not respond to Party 2's submission of materials for its review for [specify period], Party 2 may proceed to publish or present these materials.

Limitations on Liability

In no event shall either party be liable to the other party under this Agreement or to any third party for special, consequential, incidental, punitive, or indirect damages, irrespective of whether such claims for damages are founded in contract, tort, warranty, operation of law, or otherwise or whether claims for such liability arise out of the performance or non-performance by such party hereunder.

Monitoring and Breach Notification

In the event of an actual or suspected security breach involving its information system(s), Party 2 will immediately notify Party 1 of the breach or suspected breach and will comply with all applicable breach notification laws. The parties agree to cooperate in any breach investigation and remedy of any such breach, including, without limitation, complying with any law concerning unauthorized access or disclosure.

Remedies in Event of Breach

The parties recognize that irreparable harm may result in the event of a breach of this Agreement. In the event of such a breach, the non-breaching party may be entitled to enjoin and restrain the other from

any continued violation. This section shall survive termination of the Agreement. In the event that a breach is identified and it is determined by the non-breaching party that (a) individual or public notification is required and (b) that the requirement for notification is substantially caused by the other party, the party responsible for the breach shall be liable for the reasonable costs incurred by the other party to meet all federal and state legal and regulatory disclosure and notification requirements, including, but not limited to, costs for investigation, attorneys' fees, risk analysis, and any required individual or public notification, fines, and mitigation activities.

Signatures
Party 1 Name, Title, Date
Party 2 Name, Title, Date

Additional sections, as appropriate

Contacts
Party 1's designated contact concerning this Agreement is Name, Title, Address, Phone, Email. Party 2's designated contact concerning this Agreement is Name, Title, Address, Phone, Email.

User Training
Party 2 will annually sign an acknowledgment that all individuals authorized to have access to disclosed data have been instructed, as specified by Party 1 in [specify], with regard to the confidential nature of the data, and that each authorized individual has taken Party 1's [specify training]. Party 2 will take all necessary steps to ensure that the individuals who have access to data comply with the limitations on data use, access, disclosure, privacy, and security set forth in this Agreement. Such steps will include, but not be limited to, requiring each individual with access to data to acknowledge in writing that he/she understands and will comply with such limitations [specify Non-Disclosure Agreement terms, as applicable].

Public Information
To promote organizational transparency, and in support of data discovery for current and future researchers, Party 2 may publish non-

sensitive data documentation to public-facing websites. This documentation may include a project abstract, description, or summary of results.

Use of Name

Neither party will use the other party's name, logos, trademarks, or other marks without that party's written consent.

Community Stakeholders

The parties agree to engage community stakeholders in the course of this research project. No confidential data will be released or discussed with third parties, but the parties may agree to disclose de-identified aggregate reports to support their initiatives and engage community stakeholders.

Costs

This project shall not result in the transfer of funds from one party to another. Party 1 agrees to provide technical assistance to Party 2 to develop and deliver the initial data extract. If the parties determine that additional staff or supports are necessary at any stage of this research project, Party 2 agrees to seek funding to support those needs.

Appendix B

Toolkits and Guides

Links to these online resources can be found in the Online Appendix at admindatahandbook.mit.edu/book/v1.0/dua.html#dua-appendix.

California Accountable Communities for Health Data-Sharing Toolkit
This toolkit is produced by the University of California Berkeley Center for Healthcare Organizational and Innovation Research and sponsored by the California Health and Human Services Agency and University of California Berkeley, School of Public Health. This report summarizes seven parameters for data sharing, Purpose/Aim, Relationship/Buy-in, Funding, Governance and Privacy, Data and Data-sharing, Technical Infrastructure, and Analytic Infrastructure while observing that parties will have varying levels of maturity and expertise across these categories.

CMS Administrative Simplification: Covered Entity Guidance
This clickable guide helps identify whether an organization or individual is a covered entity under the Administrative Simplification provisions of HIPAA. It is a good example of a straightforward tool that aids decision-makers to understand what laws apply to whom.

Department of Education Data Sharing Tool for Communities
This toolkit is designed to simplify the complex concepts of FERPA. It covers three primary focus areas: understanding the importance of data collection and sharing, understanding how to best protect student privacy when collectively using personally identifiable information from students' education records that are protected by FERPA, and understanding how to manage shared data using integrated data systems. It includes a sample MOU and sample consent form.

Health Care Systems Research Network DUA Toolkit
This toolkit includes a useful flowchart called "When do I need

a DUA?" and a good glossary of terms, especially for health or healthcare projects.

National Association of County & City Health Officials (NACCHO) Data Sharing Framework
This report titled "Connecting the Dots: A Data Sharing Framework for the Local Public Health System" focuses on DUA content areas needed by local public health officials. It includes a case study involving data access in a Colorado community.

National Governors Association, Improving Human Services Programs and Outcomes Through Shared Data
More for policymakers than practitioners, this brief includes short examples of how data sharing helped states and their residents in Indiana, Kentucky, Maryland, Massachusetts, North Carolina, Pennsylvania, Rhode Island, South Carolina, Virginia, and Washington.

National League of Cities Sharing Data for Better Results Guide
Prepared with Stewards of Change, this guide was written for officials, agency leadership and managers. It highlights their incentives to share data, what information can be shared, and who can receive the information with specific examples across domains including education, health, mental health, substance abuse, human services, and criminal justice. They include sample MOUs from two counties, a city, and a state and have an appendix listing major federal laws and regulations.

Sharing Data for Social Impact: Guidebook to Establishing Responsible Governance Practices
Produced by Natalie Evans Harris, a program fellow with the Beeck Center for Social Impact and Innovation, this guide is for those who take action on the data and drive impact. The guide focuses on three phases: building the collective, defining the operations, and driving impact.

Agreement Collections

NNIP's Collection of Example Data-Sharing Agreements
This collection of agreements comes from multiple domains including

labor and human services, department of motor vehicles, criminal justice, education, housing, and health and healthcare. It also includes some generic agreements and other materials, such as an information security incident protocol, breach plan, and sample confidentiality pledge.

Data2Health Data Use Agreement Library
An analysis of DUA practices across 48 Clinical and Translational Science Award (CTSA) institutions, this collection includes DUA templates, forms to request DUAs, and policies and guidance documents.

Drexel Data Sharing Agreement Repository (DataSAR)
This repository is a collection of DUAs, samples, contracts, use policies, and forms. It can be filtered by domain and discipline. This collection is aligned with Drexel's Licensing Model and Ecosystem for Data Sharing Initiative.

Contracts for Data Collaboration
This collection contains DUAs for domestic and international government administrative data and private sector information. The site also includes a guide describing forms of collaboration and explains how they categorized DUAs based on Who, What, When, Where, Why, and How the data sharing was occurring.

Administrative Data Research Initiative Data Sharing Index
This index, a collection of standards, guides, and templates, is searchable by geographic categories including city, county, state, or federal and domain categories such as education, health, housing, human services, justice, or workforce.